DOCUMENT RESUME

ED 044 478 UD 011 055

TITLE	Do Teachers Make a Difference: A Report on Recent Research on Pupil Achievement.
INSTITUTION	Office of Education (DHEW), Washington, D.C. Bureau of Educational Personnel Development.
REPORT NO	OE-58042
PUB DATE	70
NOTE	186p.
AVAILABLE FROM	Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 (Catalog no. HE-5.258:58042, $0.75)
EDRS PRICE	EDRS Price MF-$0.75 HC Not Available from EDRS.
DESCRIPTORS	*Academic Achievement, *Cognitive Development, Effective Teaching, School Organization, School Personnel, *School Role, *Teacher Attitudes, Teacher Behavior, Teacher Improvement, Teacher Motivation

ABSTRACT

This collection of essays concerning recent research on pupil achievement focuses on the role of teachers. The papers served as the basis of discussions during a day-long conference in February, 1970, at the Office of Education. Topics included models of school effectiveness, teacher quality, teacher attitudes, and policy implications. While the state of research on the effects of teachers on pupil achievement is considered still primitive, a few tentative indicators are held to be emerging. From the papers in this collection, one is led to believe that schools can and do make a difference in the development of youth. Beyond this, it is thought that teachers are the single most important element in the school. The public policy implication is that more available resources must be devoted to the development of methods for recruiting, preparing, and utilizing quality educational personnel. It is held that the fact that great numbers of children are not learning to read and are not receiving other basic tools essential for productive living demands that ways to make teachers, administrators, and all educational personnel more effective be found. (Author/JW)

ED0 44478

OE-58042

DO TEACHERS MAKE A DIFFERENCE?

A Report on Recent Research
on Pupil Achievement

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

UD011055

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Office of Education
Bureau of Educational Personnel Development

Superintendent of Documents Catalog No. HE 5.258:58042

U. S. GOVERNMENT PRINTING OFFICE
WASHINGTON : 1970

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402 - Price 75 cents

FOREWORD

In a statement of goals developed in late 1969, U.S. Education Commissioner James E. Allen, Jr., said that the Office of Education must add a strong, determined advocacy of needed reform and improvement to its traditional service responsibilities. I can think of no area where it is more urgent to exercise such leadership than the subject of these papers—the relationship between student performance and teacher performance.

There is little agreement regarding the locus of the problem of school failure. We are not, however, without theories—some based on research and others on intuition.

Intuitively we know that teachers do make a difference—both positive and negative—in how a student performs, in his level of achievement, in his behavior, in the values he acquires. If teachers did not make a difference we would be satisfied with schools run and operated wholly by machines.

One problem we face when we try to measure teacher performance is that we evaluate statistics when we should be evaluating human relationships. Few would doubt that individual teachers do have a tremendous influence on individual children. We cannot evaluate a teacher's competency in teaching reading or math without also evaluating his ability to interact with the child he is teaching. A child's basic needs go beyond reading and math. They include the need for dignity and respect for another human being whom he can trust. A teacher who cannot meet human needs is not likely to meet educational needs.

Those who say schools and teachers are of little or no consequence in the educational process have an obligation to offer an alternative to the current system. In the absence of such an alternative, we in the Office of Education have an obligation to do everything in our power to see to it that schools and teachers are a *positive* influence.

In an effort to learn how we can do this more effectively, we invited a select group of educational researchers to prepare the papers which follow. These papers served as the basis of discussions during a day-long conference in February 1970 at the Office of Education. They illustrate the best of recent research on the factors which influence pupil achievement. Obviously, the views expressed in these papers are those of the authors and do not reflect official policies of the U.S. Office of Education. And while the state of the research art in this field admittedly is still primitive, a few tentative indicators are beginning to emerge. These indicators have significant public policy implications.

The research reported in this publication leads us to believe that, contrary to some earlier indications, schools can and do make a difference in the development of youth. Beyond this, it is clear that teachers are the single most important element in the school—more important than the quality of facilities, the quantity of equipment and materials, or the level of financing.

The public policy implication is clear. We must devote more of our available resources to developing improved means of recruiting, preparing, and utilizing quality educational personnel. The fact that great numbers of children are not learning to read and are not receiving other basic tools essential for productive living demands that we find ways to make teachers, and administrators, and all educational personnel more effective. This we intend to do. This was the intent of the Congress when it passed the Education Professions Development Act. This is the function of the Bureau of Educational Personnel Development.

The Bureau is putting money and energy into programs designed to recruit and train educational personnel who will be effective. It is working to bring about change in the institutions responsible for training teachers and teachers of teachers. It is searching for ways to reorganize the teacher's time so that productive teachers will have opportunities to function productively. And it is developing the means of evaluating all of these endeavors on the basis of pupil performance.

Our goal, of course, is to find more efficient ways to deliver all educational services at all levels. The research indicates that to do so we must first improve the quality of teaching.

Finally, I wish to acknowledge the contributions of Mrs. Iris Garfield, Director of the Division of Assessment and Coordination, and two members of her staff, Mr. Peter A. Hartman and Mrs. Patricia Wagner, in arranging the conference for which these papers were prepared and in preparing these papers for publication.

Don Davies
Associate Commissioner
Educational Personnel Development

CONTENTS

Page

Foreword iii

Chapter 1
Do Teachers Make A Difference? 1
Alexander M. Mood

Chapter 2
A Survey of School Effectiveness Studies 25
James W. Guthrie

Chapter 3
A New Model of School Effectiveness 55
Henry M. Levin

Chapter 4
The Production of Education, Teacher Quality, and Efficiency 79
Eric Hanushek

Chapter 5
Teacher Attributes and School Achievement 100
George W. Mayeske

Chapter 6
The Association of Teacher Resourceness with Children's Characteristics 120
Stephan Michelson

Chapter 7
Policy Implications and Future Research: A Response . . . 169
Robert M. Gagné

Chapter 8
Comments on Conference 174
James S. Coleman

Appendixes 176

Chapter 1

DO TEACHERS MAKE A DIFFERENCE?

Alexander M. Mood

This volume brings together some of the current outstanding analytical work concerned with appraising teacher effectiveness. Besides several original papers there is an extensive survey by James Guthrie, George Kleindorfer, Henry Levin, and Robert Stout of a number of recent illuminating quantitative studies. My overview of the conference that generated these papers will not abstract them but will attempt to present a fair answer to the two major questions to which it was directed. On the one hand it was intended to bring us up to date on what we can say with some assurance about the effectiveness of teachers. Its second objective was to give some direction as to what we might do next to improve our understanding of how teachers are effective and, by implication, to help teachers increase their effectiveness.

The *we* in these sentences actually refers only to myself, but I hope it is not seriously unrepresentative of us participants in the conference, or most of us educators or sometimes even us citizens of the United States. There is a third and final section of the overview which presents some thoughts about how trends of the times may change teaching; these are purely personal speculations which have no connection with the conference or the views expressed by the participants.

What Does Analysis of Data Tell Us?

Many of the important analyses use data gathered by the U.S. Office of Education in its 1965 Equality of Educational Opportunity Survey of the U.S. public schools. It has often been called the Coleman Survey after James Coleman who had the major responsibility for carrying it out but in deference to his desire that the contributions of others not be slighted we shall refer to it simply as the EEO Survey. The Survey went further than any previous one had in attempting to gather information about the whole complex of factors affecting childrens' education; in addition to data about childrens' achievement there was information about their socioeconomic status as well as about some of the education-related attributes of their parents; besides school and teacher data there was information about communities in which the schools were located.

With respect to teachers there were conventional data about teachers' age, sex, race, socioeconomic status, education, experience, certification, salary, and professional activity. There were also items which attempted to get some indication of the quality of the institution where the teacher was trained, of teacher attitudes toward minority groups, and of teacher morale. In the analysis of the data not any of these indicators turned out to be a particularly powerful discriminator for predicting student achievement but most investigators find that socioeconomic status, education, experience, and salary have statistically significant correlations with achievement in the expected direction. The item that seems to discriminate best is the teacher's score on a brief self-administered test of verbal facility. The test consists of a list of 30 sentences—each having one word missing and each having a list of five words from which one was to be selected as the most logical selection for the missing word. Hanushek, Levin, and Michelson all find it to be the most useful explanatory variable. Referring, for example, to Hanushek's table 1 we observe that its elasticity is four to six times as large as that of teacher experience. That is, the regression equation connecting these two variables to achievement indicates that a percentage increase in teacher verbal score is far more effective than an equivalent percentage increase in experience in increasing student achievement. This particular finding would not be of great practical interest if it should turn out that verbal score was a far more expensive commodity than experience. Levin, in a previous paper on related research, took the next step and priced these things out to show that verbal score is not especially expensive. (See Levin, in References.) This kind of cost analysis is something that everyone agrees must be done but rarely does one ever do it. Let us hope that Levin's example will encourage all of us to pay more attention to the important task of relating research results to the real world.

Having raised that issue we must point out that not much attention can be paid at present to the size of coefficients in regression equations or structural equations. A time will come when they will be extremely valuable but the state of model development in education is so primitive today that we do not even have a satisfactory set of variables. Thus, verbal ability is a proxy for a number of important attributes of a complicated entity called a teacher. If we went about increasing the verbal ability of teachers, the increase that might result in student achievement would be far less than what would be calculated by using the equation that relates it to achievement. The reason is that a specific increase in verbal ability would probably not be accompanied by a corresponding increase in all the other attributes that verbal ability is serving as a proxy for.

This point might be a little clearer if we think of the variable "reading matter in the home," which has a significant coefficient in any regression equation relating achievement to home background. A heavily weighted item in that variable is "presence of a dictionary in the home." If one seriously believed the regression coefficient he would rush out and buy a dictionary for every home that did not have one; he could thereby expect to bring about a huge nationwide increase in achievement at trivial cost. Of course the increase would not materialize because the dictionary is actually a proxy for a number of other educationally efficacious properties of the home which would not magically appear with the addition of a dictionary. A great deal of fundamental development work will have to be done before we can have any confidence that we have a reasonably complete set of variables suitable for the educational model; only then can we begin to believe the calculations based on coefficients in equations and begin to make the policy recommendations implied by them. *Until our models become a great deal more sophisticated they will be of very limited use to policymakers and administrators.* Michelson's paper has an excellent discussion of these problems.

Both Hanushek and Levin point out the substantial implications for personnel policy that follow from the fact that a simple performance indicator (verbal ability) seems to be so superior for judging the quality of a teacher to the indicators commonly used by educational administration (certification, experience, amount of graduate work, and advance degrees); certainly a very serious question is raised about the incentive system in education if salary (which is based upon the common indicators) discriminates achievement scores weakly. In any case, the conference participants agreed that the available data convince them that *teacher performance indicators are more relevant for judging teacher effectiveness than certification, education, and experience.* This conclusion should surprise no one; it has long been one of the basic tenets of personnel administration in the commercial world; there, rewards are based almost entirely on results and almost not at all on credentials (beginners excepted).

Does salary discriminate weakly? We think so despite the fact that when one relates student achievement scores to teacher salary directly in a simple regression they are usually found to be closely associated; that is, salary seems to discriminate rather well. If one adjusts achievement scores to account for the socioeconomic status of the children, then there is almost no relation between the adjusted scores and salary. We are at a dilemma which will plague us throughout our examination of the statistical evidence. The evidence is much too rudimentary to give us definite answers. We are just barely beginning to construct a quantitative framework for getting at these questions. It will be quite a long time before we get reliable quantitative guidance from it. All we can say about this matter at the present time is the following: children from well-to-do, well-educated families tend to get higher achievement scores; children having higher salaried teachers tend to get higher achievement scores; higher salaried teachers tend to be found in well-to-do school districts; there is insufficient evidence to determine how much of the higher achievement should be attributed to the home and how much to the teachers.

These same observations apply as well to other teacher characteristics. Thus, with respect to experience, experienced teachers develop seniority and hence some choice about where they teach; they tend to gravitate to the comfortable suburbs; hence one finds good association between student achievement and teacher experience. How much of the higher achievement should be attributed to teacher experience? The present rudimentary state of our knowledge permits us to make no reasonable estimate of it.

This basic difficulty with the existing quantitative knowledge of the educational process is consistently brought out by every investigator. Student achievement correlates with almost any school attribute and it is no trick to build up a set of attributes which will generate a sizable correlation. The same can be done with home attributes or with community attributes. When one tries to control on one set in order to assess the effect of another set he finds that he has overcontrolled and the sought effect is very small—vastly smaller than it would have been without the control. Thus the original report on the EEO Survey regularly found extremely small school effects of any kind after adjustment for students' socioeconomic status had been made. Several of the studies surveyed in Guthrie's paper exhibit the same phenomenon; sometimes school effects are found to be statistically significant even after adjustment for student socioeconomic status but they are nevertheless quite small and the significance is more a result of large sample size than of real magnitude. We may conclude as a general result of these findings that *teacher effects will be seriously underestimated if achievement data are first calibrated for student socioeconomic status.* We cannot actually demonstrate the truth of that statement because we are not able to estimate teacher effects in isolation but most investigators are convinced the statement is true.

Mayeske's paper deals with these difficulties in a quantitative way by focusing on reductions in variance rather than on regression coefficients. This was the primary analytical technique used in the original analysis of the EEO Survey data (Coleman, et al 1966), but in Mayeske's paper it has meanwhile become a considerably more powerful tool and in addition it has been applied with a great deal more care and sophistication than was possible in the original analysis (which was pushed by various delays in getting the data too close to the Congressional deadline for submitting the report).

For the benefit of those not familiar with statistical methods I shall take a paragraph to indicate roughly what Mayeske's analysis does. Different ninth grade children have different achievement scores for many reasons: differing abilities, differing parents' education and interest in schooling, differing abilities of their teachers, differing interests themselves, how they felt on the day of the test, and so on. Statisticians calculate an index of the extent to which the scores jump around; it is called the variance (and calculated by subtracting the average score from each score, squaring those differences, adding the squares together, and dividing by the number of scores; that is, it is the average of the squares of the differences). If the scores are first adjusted for parents' education, then the variance of the resulting adjusted scores will be smaller; let us suppose for illustration that the adjustment reduces the original variance by 25 percent. Now let us consider a second adjustment using, say, teachers' verbal ability instead of parents' education and suppose that that adjustment reduces the original variance by 20 percent. Finally let us adjust the scores for both parents' education and for teachers' verbal ability and suppose, for purposes of illustration, that the double adjustment reduces the original variance by 35 percent. The results of this set of calculations are described thus: of the combined reduction in variance of 35 percent, *10 percent is uniquely associated with teachers' verbal ability* (because that, in the combined adjustment, reduced variance 10 percent over the 25 percent achieved by the parents' education adjustment alone); *15* *percent is uniquely associated with parents' education* (because that, in the combined adjustment, reduced variance 15 percent over the 20 percent achieved by the teachers' verbal ability adjustment alone); and *the remaining 10 percent (35 percent minus the two unique parts) is common to both parents' education and teachers' verbal score.* There is no way to tell whether that common 10 percent should be attributed to parents or to teachers or whether it should be divided between them somehow.

The numbers in the above paragraph were purely hypothetical. Some actual numbers may be found in Mayeske's table 1 which illustrates especially well the extraordinary amount of overlap between home and school attributes. The table refers to two sets of variables (instead of just two variables as in the above paragraph); one set called B refers to the students' background and the other set S refers to attributes of the school. The table shows that of the total reduction in variance of a set of scores (this table refers to the reduction, not the whole variance, so the total reduction is called 100 percent) achieved by the B and S sets in combination, 94 percent of the reduction can be accomplished by the B set alone and 88 percent of the reduction can be accomplished by the S set alone. The overlap (or commonality) of the two sets is 82 percent which is quite a large number relative to the two unique parts; it indicates that the B set is a very poor set of variables for getting specifically at background effects and that the S set is a very poor set of variables for getting specifically at school effects. If the scores are adjusted first by the B set, 82 percent of the 88 percent that the S set could have removed by itself will have been removed by the adjustment and only 6 percent will remain to be identified with the S set. This and the other results presented by Mayeske make it clear to all investigators that *the present rudimentary state of our quantitative models does not permit us to disentangle the effects of home, school, and peers on students' achievement.*

The commonality model has the advantage over the linear equation models of not encouraging people to substitute numbers into equations and then believing the resulting calculations. The size of commonalities supplies us a good criterion for the degree of primitiveness of our models; the smaller the commonalities get, the more confidence we can have that our variables are actually measuring the things we are trying to measure. When we can get those commonalities down to perhaps half their present size or smaller, we can joyfully abandon the commonality model and move to the much more illuminating regression models and still more illuminating structural models that have been described in the papers of Michelson and Levin.

Can commonalities be substantially reduced? Can home, school, and peer effects be disentangled? Probably not entirely but surely to a considerable degree. The problem at present is that our measures are far too crude. We are using simple items that are really only proxies for the items we should be measuring. Hanushek points out clearly in his Ph.D. dissertation that many of the items that go into socioeconomic status are simply evidences of income. Family income does not teach children. We have to get at what parents do that helps their children learn; we shall doubtless find that many parents without much income do those things too and that their children consequently tend to do well at school. It is also fairly obvious that we have extremely crude measures of teacher quality and I shall explore that consideration further in the next section. The simple proxy devices have the unfortunate property that, for example, they can represent community or parent or teacher attributes even though they were meant to measure student attributes. It is no wonder that we are having great difficulty getting any real grip on teacher effect.

We can only make the not very useful observation that *at the present moment we cannot make any sort of meaningful quantitative estimate of the effect of teachers on student achievement.* Many investigators believe that teachers may be the most important factor in educational achievement for most children and are at worst second only to parents. That belief rests largely on judgment and it may well be true; unfortunately it does not give us any clue as to how it operates and without that it is not of much use to policy formulation or administrative practice.

What Must We Find Out?

If, as has been said of investigations in the physical sciences, the mark of a successful experiment is the number of fundamental questions it raises, then the EEO Survey was quite a success. It was an attempt to obtain some sort of comprehensive quantitative understanding of the whole range of basic factors that enter into educational achievement. We did not get much fundamental understanding out of it but we did get some real sharpening of fundamental questions. Now we can see that the measuring instruments were altogether too crude (except for the tests which measured academic achievement). They were crude because they did not begin to cover all the important facets of such complex factors as parents, teachers, and peers; not only were they impossibly brief, they relied too much on easy to get but not very discriminating proxies. The result is that we have only the barest beginning of quantitative comprehension.

So we must try again and keep trying and improving and refining. We absolutely must pin down the connections between the inputs and the outputs of education; without that kind of theoretical structure we can flounder indefinitely in our efforts to improve the process.

One set of inputs to the process consists of youths with various levels of intellectual and behavioral competence. Another set of inputs consists of teachers with various competences. There are other inputs. The outputs are youths with higher levels of competence (and incidentally teachers with greater experience). Very broadly speaking, the competences which education is intended to develop in students are of two kinds. There are skills and knowledge in such areas as:

Communications,
Mathematics and computer languages,
Natural sciences,
Social sciences,
Humanities, and
Arts;

and there are matters of personal development such as:

Social competence,
Responsibility,
Self-confidence,
Creativeness,
Ethics, and
Carefully thought out personal goals.

We have reasonably good instruments for measuring skills and knowledge; we have essentially no capability at all when it comes to measuring the aspects of personal development. Merely to quantify the outputs, therefore, we must carry out a substantial instrument development program which will be largely in the realms of psychology and belief rather than in the conventional academic realm. Only then can we begin to explore how these personal development outputs change as teacher and other inputs change in the manner that the papers included in this volume are beginning to do with respect to academic outputs.

I have written elsewhere (in a paper included in the bibliography) of how a comprehensive analytical model can be developed which will unify explorations of this kind and form a basis on which can be built a verifiable body of knowledge about the operation of the educational system. A very similar model is presented in the first part of Hanushek's paper; more sophisticated structural models are presented in Michelson's and Levin's papers. This kind of theoretical knowledge is essential to formulation of effective educational policy and to effective management of school systems. We see in Levin's paper an excellent illustration of the kind of policy guidance that could flow in quantity from a valid quantitative model of the system.

The major inputs to the model besides youths and teachers are parental inputs, peer inputs, community inputs, inputs of the larger society, school administration, curriculum, and school facilities. Since we are primarily concerned here with teachers, it may be worthwhile to elaborate that particular input in order to see how far we have yet to go before we can have any confidence that we are able to assess teacher-pupil interactions. I am not speaking of understanding the interactions; I am speaking merely of assessing their effects in terms of educational accomplishment. That is, as several investigators of the EEO Survey data have found, the verbal ability of the teacher is definitely associated with pupil achievement. We do not need to go into the question of how the ability operates to increase achievement; one can make more or less reasonable speculations about it but those are not essential to the construction of the model or to policy utilization of the model. It is sufficient that we can measure achievement, that we can measure verbal ability, that we can estimate the degree of their association, that we can demonstrate it by experiment, and that any objective investigator would come to essentially the same conclusions if he should attempt to duplicate the analysis and the experiment.

We must develop a comprehensive model for this scientific purpose itself as well as for policy and management purposes. Experimental results cannot be duplicated without it. Education is such a complicated endeavor that it is really impossible to duplicate experiments faithfully; for one thing teachers and pupils cannot be duplicated. Experimenters can only do the best they can to carry out approximate duplication; then they must adjust their results to take account of the deviations of experimental conditions from true duplication. *The model enables such adjustments to be made. Until we have one, there will be no operationally effective science of educational systems* because there cannot be a science without a means for determining what is and what is not duplicatable.

What must be measured about teachers? Every attribute that is significant to teaching effectiveness or, as Robert Gagne says, is significant to the ability of teachers to facilitate learning. Many of us are convinced that verbal ability (accurate understanding of the meaning of words) is one. There may be 50 others—more or less. The sole source of that number is the fact that I have taken a little time to try to list teacher attributes that might conceivably be as important to learning as understanding the meaning of words. The list follows, arbitrarily classified under five headings.

Dedication to the Educability of all Children

- Conscientiousness
- Humaneness
- Patience
- Sensitivity
- Optimism
- Tolerance
- Responsibility
- Fairness
- Inclination to praise success
- Inclination to react to mistakes with reassurance

Ability to Communicate

- Verbal ability
- Fluency
- Lucidity (in the vocabulary of the students)
- Poise
- Sincerity
- Tact
- Expressiveness
- Good humor
- Adaptability
- Tendency to use illustrations and examples

Ability to Motivate

- Empathy
- Enthusiasm
- Helpfulness
- Resoluteness
- Persuasiveness
- Friendliness
- Earnestness
- Generosity
- Open-mindedness
- Charm

Ability to Organize and Manage a Class

- Leadership
- Confidence

Maturity
Common sense
Intellectual honesty
Responsiveness
Realism
Integrity
Equanimity
Attentiveness
Capacity to appraise and evaluate

Ability to Create Learning Experiences

Capacity to diagnose and analyze learning difficulties
Familiarity with teaching methods
Tendency to experiment
Originality
Resourcefulness
Curiosity
Artistic ability (particularly to draw illuminating pictures and diagrams)
Imaginativeness
Ability to dramatize

There is a sixth important classification having to do with the teacher's knowledge of a chosen field in which to teach but we shall omit consideration of that because instruments for measuring those attributes already have a long history of development and are in a reasonably satisfactory state.

The listed attributes doubtless overlap to a considerable degree; the projected model will require that the overlaps be determined and that the list be pruned down in order to eliminate any near duplicates. That is necessary to prevent collinearities from injecting instability into the model. It will require a large investigation. I am reasonably certain that we shall get essentially nowhere by trying to make do with combinations of existing personality tests such as, for example, the Minnesota Multiphasic. We shall simply have to sit down and do the slow laborious work of devising a list of a dozen or so questionnaire (or interview) items for each and every one of these teacher attributes—items thoughtfully and narrowly directed specifically to the attribute. Then a large sample of data must be obtained from teachers and factor analyzed by the same procedures that Mayeske and his colleagues used in developing their indices for the EEO Survey data. While this kind of sweeping attack on the dimensions of teacher effectiveness will not guarantee that every dimension will be uncovered, perhaps most investigators will feel reasonably confident that no important one has been omitted altogether; these imprecise attributes do overlap and it is likely that any others that might be measured will overlap these to some extent and hence will be represented by these to that extent.

Once this analysis has been carried out then construction of the next stage of the model can begin. That stage will resemble the relations we see in the papers of Hanushek, Levin, and Michelson which connect student achievement to teacher characteristics. The difference will be that something approaching the full force of teacher effect will be represented. (My personal belief is that it has been dreadfully underrepresented in all studies that have been carried out thus far; that is, that there are many important dimensions of teacher quality that have insignificant overlap with the dimensions we have been accustomed to measure.) Full representation will give us real potential for assessing the whole teacher effect, for better differentiating home and school effects, and for determining the relative importance of the various teacher attributes. This last information will give crucial policy guidance for teacher education and for counseling those who are considering preparing for teaching as a profession.

Another very important matter discussed by Michelson can then be explored to the probable great benefit of school administration. That has to do with the variety of students and the likelihood that different kinds of students will learn best with different kinds of teachers. Some teachers just naturally turn some kids off. Learning depends so strongly on teacher-student interactions that there must be considerable potential for improvement of the educational process by developing procedures for assigning students to teachers in a way that will enhance those interactions.

In order to make valid connections between student achievement and teacher characteristics it is essential that differential student achievement be associated with specific teachers (Hanushek and Michelson). That is, the students must be measured at the beginning of the school year and again at the end of the school year. The analysis of teacher effects must use the gains in achievement levels—not the achievement levels themselves.

The quality of this proposed model development program will depend very much on our having instruments for measuring student achievement in personal development as well as for measuring academic achievement. Teacher attributes important for the former may well be somewhat different from those that are effective for the latter. It would be an inexcusable blunder to depreciate the qualities of those teachers who are doing an outstanding job of personal development of students.

There will apparently be some difficulty about associating personal development increments with specific teachers in secondary schools because students have several teachers. In the elementary grades where students normally have a single teacher the difficulty will not arise (as Hanushek observes). But even in secondary schools the difficulty may be more apparent than real. Every student is exposed to a set of teacher attributes (in the language of the model); in elementary schools that set for a particular student happens to correspond to a single teacher; in secondary schools the set for a particular student consists (to a first approximation) of the same attributes averaged over the teachers whose classes he attends. The main difference might be that the secondary school student will be less subject to extreme values of an attribute and hence a larger sample of data will be necessary to determine how a specific student personal development outcome is associated with a given teacher attribute.

How May Teaching Change in the Future?

The purpose of an overview is not only to consolidate present knowledge but to use it to deduce plausible directions for the future. The preceding considerations naturally lead me to hope that the future of educational research includes a massive exploration of the connections between teacher-student interactions and learning. Considering the kinds of interest that have developed at this conference perhaps it is not a wholly hopeless hope. Many able analysts are anxious to work on these problems. The work is an absolutely essential prerequisite to any substantial improvement of the educational process. Only the resources are lacking to get it under way and I am sure those at the conference who represented the U.S. Office of Education are working diligently on that matter.

One of the conference participants, Professor Doxey Wilkerson of Yeshiva, correctly pointed out toward the end of the conference that exactly nothing had been said or written about how teachers make a difference. The conference produced no suggestions for teachers or for teachers of teachers. I shall take it upon myself in the remainder of this overview to make a small gesture toward repairing that omission. It should be noted, though, that the conference was not much directed to that question despite its title; its primary aim was to discover the extent to which hard data could be used to estimate how much difference teachers do make.

In any case paucity of solid information about the relation of teaching to the learning process will naturally force many of us in education to look more attentively than we might otherwise to indirect information that may help us understand teaching and how it may develop over the next several years. We cannot escape indulging in a great deal of speculation in this endeavor but on the other hand it is essential that someone construct some conception of teaching of the future so that young persons planning to become teachers will have a glimpse of the various roles they might fill and so that those who are teaching teachers will have some clues as to how their activities may change. So I make no apology for generalizing as best I can about the implications of whatever signals I am able to detect.

Theater Arts

A number of clues point to the likelihood that acting, directing, dramatic writing, animation, and staging may become an essential part of teaching. A great many teachers may be doing nothing else; they are the ones who would be teaching huge unseen classes via films and TV programs.

I realize that it is not fashionable just now to get excited about the wonders of technology and I agree with many of the criticisms of it. The idiot box will never replace the teacher. The impersonality of the box is a staggering liability in the age of increasing urbanization which puts increasing reliance on practiced social intercourse. The box cannot notice that it has lost the child; it cannot hear his questions; it could not answer them anyway. Worst of all it cannot bend even slightly to the child's desire that it deviate from its program. (Some programs have considerable built-in flexibility; I am referring to excursions outside that range of flexibility.) Nevertheless there is one thing it does exceedingly well and that is transmit information at great speed. A picture is worth a thousand words and furthermore it can be grasped in about the same amount of time as can one word. It is an undeniable fact of physics and physiology that nothing else can begin to approach colored pictures for transmitting large numbers of bits of information per second to the human brain. That fact has a large contribution to make to educational effectiveness. We cannot give it much time during the school day but while it is operating it can be a powerful tool.

The box can do other things. It can be an infinitely patient drillmaster. And despite its impersonality, we have all seen in good movies how accurately it can present deep human emotions and complicated human behavior with an indelibility that words could never match. These boxes will blossom in the hands of teachers skilled in using them and supplied with material created by teachers skilled in preparing them. So much for boxes.

You can lead a child to Chaucer but you can't make him think. (Sorry 'bout that.) Showmanship is not only for teachers who are creating fascinating educational materials. Showmanship is for all teachers. There was a time, now long past, when school may have been something of a relief to children burdened with arduous chores at home or on the farm. Nowadays they mostly watch television at home. In comparison with that, school is usually a drag strictly from dullsville.

It will not cease being a drag until we start fighting fire with fire. A humdrum performance simply will not hold the attention of our children; they will switch to another channel—leaving education to drone on to the other. Unfortunately, the marijuana channel seems to be sort of interesting.

Student Participation

It appears to me to be reasonable speculation that teachers of the future may make a large difference by fully including students in all aspects of carrying out the educational enterprise. This will require revolutionary changes in organization, schedules, and curriculums. At present, the organizational arrangement of teachers and pupils in a school is almost everywhere determined by the simple venerable concept of dividing the pupils about equally into as many groups as there are teachers and then placing each group in a room with one teacher.

It will not be easy to change because it is established by long tradition and is therefore buttressed by the expectations of teachers, children, and parents; by the existing administrative structure and hence the whole experience of school administrators; by the training of teachers; by the design of school buildings; by the pattern of all the tools available to teachers; by a salary structure that awards the best teacher the same wage as the poorest teacher with the same training and experience; and most of all, by the budget which unmistakably spells out the pupil-teacher ratio.

Nevertheless in recent years a number of ideas have been put forward for changing the traditional pattern; some of them have been given limited trials with considerable success. One is the team teaching arrangement which puts two or more teachers in a classroom for certain special instructional purposes. Another contemplates putting layers of organizational structure into the teaching staff so that the more able teachers supervise the younger or less able teachers in various ways. Another would add still more echelons to an organizational structure for teachers by including paraprofessionals and teachers' aides in the school staff. Another would attempt to introduce great variability into class size so that a better match might be made between intensity of instruction and the difficulty of curriculum material. Another would rotate teachers so that the best ones would teach the most difficult material. Another would use the better educated parents or retired persons or some of the older and brighter children as tutors for those children having special learning difficulties.

None of these ideas quite gets to the heart of full student participation; that requires the interweaving of teachers and pupils into a unified organization. The students must be integral elements of the organizational enterprise—not merely a group of outsiders that the organization deals with. To this end all children must regularly be assigned teaching roles. Even third or fourth grade children would spend a little time helping individual first or second grade children. As children move up through the grades, increasingly more of their time would be devoted to teaching and the size of the group taught would increase slowly.

One expected benefit of the rotation of all children through teaching roles would be enhancement of their understanding and hence identification with the goals of the school. It occurs now mainly in the interscholastic athletic programs where the staff and the students are in good agreement about the goals and therefore jointly pursue them in a productive spirit of collaboration.

Another benefit of the rotation through teaching roles would be acquisition of extensive experience in performing supervisory and subordinate roles with a wide variety of personality types. These are the roles that all students must learn well if they are to be prepared for an ever more highly organized adult society.

An additional benefit to be expected of the rotation through teacher-pupil roles is partial fulfillment of the requirement that schools provide a rich variety of social experience to assist the development of social skills. It is a critical defect of current school organization that children get hour after hour, day after day, year after year, one utterly monotonous social experience in the classroom.

The teaching experience of students should surely increase rapport between teachers and students because students will discover what a difficult art teaching is; they may have better tolerance of the shortcomings of teachers and far better appreciation of good teaching.

Student teachers will rapidly learn the disaster of being unprepared. It is one thing to shrug off failure to do one's homework among one's peers but quite another thing in front of an expectant group of younger children. Mary Kohler's Youth Tutoring Youth Program has shown that this phenomenon gives schools a powerful new dimension of teaching; when a student has difficulty with an idea, give him the task of teaching it to a couple of younger children and he will pore over it mightily.

Pedagogy, educational psychology, and individual psychology would become a significant part of the elementary and secondary curriculum. The considerations here are that: (1) the student teaching must be as effective as possible, (2) education is more and more becoming lifelong as technology accelerates and much of it will necessarily take place on the job and in the home so that all of us will be continually teachers and learners, (3) recent realization of the tremendous importance of training and education during the first 5 years of a child's life implies that all students must be taught to lead their own children effectively through those first years, (4) recent realization that the primary cause of adult failure is not incompetence but possession of annoying personality traits and the prospect that understanding of psychology by oneself and one's peers at an early age may tend to minimize solidification of such traits. Most importantly, knowledge of pedagogy and educational psychology will enable students to understand the methods and tactics that the adult teachers are using in their teaching. They will then be able to exert real intellectual influence on the educational process; there will be opened up to them a whole spectrum of reactions to the system instead of just the two available to them now (acceptance or rejection); they may even be able to force some modernization and relevance into the curriculum.

Sensitivity

A whole new conception is developing of what constitutes civilized behavior. It is a substantially lovelier and kinder concept than we have been accustomed to but it is somewhat difficult to recognize because it is usually advanced by nonestablishment young people whose behavior appears to be atrocious. It is not really atrocious but there are moments when they become outraged at what they consider to be uncivilized behavior. Those are the moments when the press puts the spotlight on them, as is perfectly natural for the press, because at those moments their behavior seems to be so inconsistent with what they are talking about. That's news. And perhaps the fact that it is news means that they may have something.

The main ingredient of the new standard of civilized behavior is the decree that psychological violence is as abhorrent as physical violence. The psychic scar is often more abominable than the scar of the lash because it keeps on hurting so long—sometimes for a lifetime. Insult, humiliation, sneer, arrogance, caste, intellectual superiority, and holier-than-thou have to go. When some of our young people experience psychological violence they react as if they had been clubbed on the head or shot in the leg; not surprisingly their reaction may be a doubled and redoubled dose of psychological violence—a dose large enough that it may have a chance to penetrate the insensitive skull of the perpetrator of the original violence.

Sensitive teachers certainly make a large difference to children. Such teachers never indulge in humiliation by design or by accident. There is no better way to keep a child ignorant than to humiliate him now and then. The humiliation rankles; every tiny facet of it demands the closest examination; try to expunge it from his mind as he may, it keeps creeping back in; obviously it cannot be displaced by such ego-insignificant trivia as the product of 6 and 9 or the spelling of Mississippi.

We are beginning to learn how to carry out sensitivity training. It would be possible for every teacher to have it. Imagine what a difference teachers may make when all of them are as sensitive as our most sensitive teachers are now. It would be hard to exaggerate the amount of additional education that might accompany that state of affairs. It is not just that unintentional teacher-created roadblocks to learning might largely disappear. That would be a very small part of it. Much more important, teachers might be better able to recognize at once when communication is failing. They might be far more expert at diagnosing students' learning problems. A whole new sympathetic mental environment could do much to erase the remaining custodial, adversarial, incarcerational vestiges of the school system. That environment might in turn generate a new level of civilized behavior on the part of the students themselves. They might become more sensitive partly as a matter of instruction but also as a result of appreciating and imitating the living example set by the teachers they encounter. Insensitivity might tend to become socially unacceptable and later unthinkable.

Philosophy of Value

It is becoming common knowledge that there is not a single unique value system; that there is not a simple rule for determining whether an act is right or wrong; that there are endless shades of gray; that some acts can be right in some quite acceptable value systems and at the same time wrong in other quite acceptable value systems; that one's personal value system cannot be identical to any other because it depends upon one's own conscience which in turn depends upon his genetic and cultural heritage. How many children have been convinced that they are utterly worthless by parents and teachers who perfidiously claim to adhere to some ridiculously stringent moral system? How many children are driven to suicide each year by that lie?

Of course parents are far more guilty than teachers but teachers are not innocent; altogether too many of them pump their quota of hot air into these adult-inflating conspiracies apparently quite unaware of the tremendous damage they may do to some children. Some misguided teachers actually appear to believe that these lies are good for children. They are not—by any stretch of the imagination. If children believe them, they are made miserable by their own behavior; if they do not believe them, they have become cynics and it is not easy to educate cynics.

The greatest benefit to developing value judgment could come from frequent thorough exploration of controversial issues. It is a most educational experience for students to hear respected authorities constructing an impenetrable case for one side of a question and another equally respected group of authorities constructing an equally impenetrable case for the other side. That is where the cultural action is. That is where society is trying to get out of some rut or other. That is how society exhibits its capacity to adapt to new conditions and to meet the future. Youths are going to live in the future. These controversies are often right in the middle of their interests. That is where relevance is. They need to understand how fragile the rational underpinnings of social institutions really are and how society actually goes about tearing them down or shoring them up.

It has been said that children are not sufficiently mature to explore such an adult matter, for example, as the recent argument between Government officials who wanted to name the TV models that start fires and the captains of the electronics industry who did not want them named. There are arguments, good and bad, on both sides. The contention that kids cannot understand and make their own evaluations of these arguments is baloney. Not only do they have excellent intuition about justice and equity, they have a great deal of sophistication. That sophistication comes from TV itself where they daily see perfectly groomed, faultlessly attired corporate executive types continually spouting in dead seriousness the utterest drivel as they peddle their sponsors' products. That drivel often includes outright lies about the marvels that flow from such products as nicotine and deodorants. If one deliberately set out to devise an educational process which would most effectively expose the shallowest and shoddiest aspects of our society to our children, he would be hard put to improve on TV as it exists today. At any rate it works; our kids know the score like no other generation of kids ever did. The United States is the greatest country in the world but there are important things wrong with it that many people believe could wreck it and our kids have a good impression of what those things are, the generation gap may save our lives; perhaps the Nation's prospects would improve if the gap were even greater; possibly we owe a vote of thanks to the racists and predatory merchants and frightened super patriots who are industriously widening it.

But other people teach kids some of the unpleasant facts of life also. I talked recently with a bright 13-year-old high school girl who had learned that in order to get an "A" in her freshman Spanish course she would be smart to sign up for German under the same teacher (who happened to owe his job to the existence of a class in German); she is not working very hard on her Spanish. "To hell with it, I can get into college with good grades in my other courses." The engaging thing about that statement is the first part; up to now she has a spotless academic record but she is not going to shed any tears that a stupid happenstance will probably bring her a "C" in Spanish. Good value judgment. The second part of her statement is not completely satisfying, is it? Reflects a little too much certainty that college is the only possible option, doesn't it?

Surely there is an acceptable value system that does not include the axiom that all able people must go to college. There are a great many careers for which college is largely a waste of time; progress along those careers might be more satisfactory if a person plunged right into them from high school and educated himself along the way in small increments as his progress required. Most business careers are in this category; so are many social service and public service careers; so are most artistic careers. Society needs able people in these careers and it is not necessary to first dump them all into the sieve for graduate schools. Let's pass over the waste of public resources spent on higher education of those for whom it does very little; maybe we are rich enough to afford it; I doubt that we are but let's pass over it. It is altogether likely that many students who do go to college cannot themselves afford the waste of 4 years and of the money that supports them.

We educators and we parents could be making a large blunder by convincing them that they are doomed to second class status if they do not incur that waste. We would be committing great numbers of blunders each year by assuring those who cannot possibly go to college that the United States has only second class status for them. We could be short changing ourselves monstrously by rating scholastic aptitude above imagination and artistic talent, and thus diverting magnificent talents away from their natural insightful creations into minor intellectual endeavors. We could be building dangerous tensions into our social fabric by labeling large numbers of people as dumb and labeling large numbers of important or necessary occupations as suitable for dumb people.

What an immense difference teachers could make by illuminating for young people the great variety of perfectly legitimate value systems! Reassurance could be brought to those who see quite clearly that their own natures are wholly incompatible with the traditional formula for success. (Whatever rung of the ladder you happen to be on, scramble frantically for the next one; when you get there scramble frantically for the next one; don't worry about where the ladder leads; it leads to the top.) The decision not to climb the ladder could be regarded as having great wisdom. Encouragement could be offered to those who are beginning halting efforts to explore other life styles and novel dimensions of personal satisfaction. Resoluteness could be imparted to those who are determined to succeed as whole human beings rather than as generators of income.

In conclusion let me repeat that I have been sifting clues and giving you my best judgment as to how teaching may make a difference—a big difference—in the future. I have been listening to young people speak and reading what they write. To the best of my ability to interpret what they are saying, I have tried to tell you where they may be taking this world. Few of us who are teachers seem to be paying enough attention to them. They are our customers and as such they are becoming more and more dissatisfied with our services; we are in trouble; the longer we stumble around in ignorance of how to do what we are trying to do the more miserable that trouble is going to make our lives.

Do teachers make a difference? Of course they do. Obviously Herbert Kohl made quite a large difference to 36 hapless children who suddenly had a fabulous stroke of luck when he walked into their classroom. There are dedicated teachers who are determined that every last child in the class will learn the material expected of him. There are uninspired teachers who are getting something across but not much. There are loving teachers who bring lifesaving affection to miserable children of acrimonious families. There are unfeeling teachers who injure children by publicly humiliating them. There are brilliant teachers who can convert a child's interest in almost anything into hard work on the very thing he needs most. There are idiots who destroy childrens' self-confidence by convincing them that they do everything wrong. There are saints who somehow civilize little demons that everyone else have given up on as hopeless. We could go on and on with statements of this kind; the point is that some teachers make

a huge difference; some teachers make a large or a medium or a small difference; a few teachers may even do more harm than good. But all teachers desire to make a big difference; they would find tremendous satisfaction in making a big difference; they could make a big difference if we would tell them how; we could if we would put some real effort into it.

References

Avorn, Jerry, *Up Against the Ivy Wall,* Atheneum, N.Y. 1968.

Bowles, S. S., and H. M. Levin, "The Determinants of Scholastic Achievement," *Journal of Human Resources,* Vol. 3, pp. 3-24, 1968.

Carmichael, S. and C. V. Hamilton, *Black Power,* Vintage, N.Y. 1967.

Clark, Kenneth, *Dark Ghetto,* Harper and Row, N.Y. 1965.

Cleaver, Eldridge, *Soul on Ice,* Dell, N.Y. 1968.

Coleman, James S., *The Adolescent Society,* Free Press, Glencoe, 1961.

Coleman, J. S., E. Q. Campbell, et al, *Equality of Educational Opportunity,* U.S. Office of Education, GPO No. FS 5.238:38001, Washington, D.C. 1966.

Conot, Robert, *Rivers of Blood, Years of Darkness,* Bantam, N.Y. 1967.

Cottle, Thomas, "Young, Creative and Trapped," *Change Magazine,* Vol. 2, pp. 20-31, 1970.

Goodman, Paul, *Growing up Absurd,* Random House, N.Y. 1960.

Hanushek, Erik, *The Education of Negroes and Whites,* Unpublished Ph.D. Dissertation, Massachusetts Institute of Technology, 1968.

Herndon, James, *The Way It Spozed To Be,* Simon and Schuster, N.Y. 1968.

Hoffman, Abbie, *Revolution for the Hell of It,* Dial, N.Y. 1968.

Hoffman, N. von, *We Are the People our Parents Warned us Against,* Quadrangle Books, Chicago, 1968.

Holt, John, *How Children Learn,* Pitman, N.Y. 1967.

Jencks, C., and D. Riesman, *The Academic Revolution,* Doubleday, N.Y. 1968.

Kershaw, J. A., and R. N. McKean, *Systems Analysis of Education,* RM-2473, The RAND Corporation, Santa Monica, California. 1959.

Kohl, Herbert, *36 Children,* New American Library, N.Y. 1967.

Kohl, Herbert, *The Open Classroom,* New York Review, N.Y. 1970.

Kunen, James, *The Strawberry Statement,* Random House, N.Y. 1968.

Levin, Henry, "A Cost-Effectiveness Analysis of Teacher Selection," *Journal of Human Resources,* Vol. V, No. 1, Winter, 1970, pp. 24-33.

Leonard, George, *Education and Ecstasy,* Delacorte Press, N.Y. 1969.

Mayeske, George et al, *A Study of Our Nation's Schools,* U.S. Office of Education, 1970.

Michael, Donald, *The Next Generation,* Random House, N.Y. 1965.

Mood, A. M., "Macro-Analysis of the American Educational System," *Operations Research,* Vol. 17, pp. 770-784, 1969.

National Commission on Resources for Youth, *Youth Tutoring Youth—It Worked,* 39 West 44th, New York, 10036, 1968.

Portman, C. and C. Weingartner, *Teaching as a Subversive Activity,* Delacorte, N.Y. 1969.

Riesman, D., N. Glazer, and R. Denney, *The Lonely Crowd,* Doubleday, N.Y. 1953.

Schwab, Joseph, *College Curriculum and Student Protest,* University of Chicago Press, 1969.

Simmons, J. L., and B. Winograd, *It's Happening,* Marc-Laird Publications, Santa Barbara, California. 1966.

Skolnik, J. H., *The Politics of Protest,* Simon and Schuster, N.Y. 1969.

Wolf, Leonard, *Voices from the Love Generation,* Little, Brown and Co., Boston, 1968.

Chapter 2

A SURVEY OF SCHOOL EFFECTIVENESS STUDIES

James W. Guthrie

In a Nation where more than a quarter of the total population is annually enrolled in schools, it borders on the heretical to contend that formal education does not or cannot make a difference in what a student learns. Nevertheless, for many interested laymen and educators, and some researchers, the so-called Coleman Report has provoked just such a heresy. Whether they gained their perception of school ineffectiveness from actually reading the Report or acquired it second hand through an interpreter or medium is a good question. Regardless, the fact remains that since publication of *Equality of Educational Opportunity*[1] the belief has become increasingly pervasive that patterns of academic performance are immutably molded by social and economic conditions outside the school. If incorrect, and if allowed to persist unexamined and unchallenged, this belief could have wildly disabling consequences. It is not at all difficult to foresee how it could become self-fulfilling; administrators and teachers believing that their school and schoolroom actions make no difference might begin to behave accordingly. Conversely, if the assertion is correct but allowed to pass unheeded, the prospect of pouring even more billions of local, State, and Federal dollars down an ineffective rathole labeled "schools" is equally unsettling.

The purpose of this paper is neither to solicit salvation for unabashed advocates of more schooling nor to grant grace to school critics and cynics. Rather, our intent is to provoke more sophisticated discussion regarding school effectiveness than has frequently been the case in the past. Our tactic in pursuing such an objective is to present a comprehensive review and analysis of school effectiveness studies, many of which have been conducted in the time since publication of *Equality of Educational Opportunity*. We begin this presentation by attempting to place contemporary assessment efforts in historical perspective. Following that, we discuss the theoretical, more accurately, "nontheoretical" nature of such studies. The remainder of the paper is concerned with a study-by-study review of recent efforts to examine systematically the impact of school variables upon student performance.

Historical Perspective

For many years, at least since public schooling became an endeavor involving many millions of dollars, laymen, educators, and researchers have been interested in making the enterprise more effective, and hopefully more efficient. This concern has been reflected in a large number of research studies dealing with school effectiveness. Early efforts were conducted for the most part by professional educators. This work is probably best characterized by the "cost-quality studies" of the late Paul R. Mort of Teachers College, Columbia University.[2] The general mode of these studies was to use per pupil expenditure levels as gross measures of the quality of a school. The "outputs" of schools were measures on a number of dimensions. In some of the better studies, the dollar inputs were related to actual measures of pupil performance. In other studies, assessment of school effects stopped short of pupil performance measures and took instead some process variable such as the rate at which the schools adopted innovative instructional practices or new curriculums.[3] The studies rather consistently concluded that those districts which spent more dollars per pupil were the most "effective," their students performed the best on test scores, attended college more frequently, etc. These findings provide a strong case for increasing school expenditures if one desires higher levels of student performance.

The simplified cost-quality studies, however, contain a serious deficiency. They do not take into sufficient account the student's capabilities prior to entry into the school or the type of experiences he participates in outside of school. In short, such studies do not control adequately for the background and environment of the pupil. What their findings tend to demonstrate is that the high expenditure districts, the Scarsdales, Grosse Pointes, and Palo Altos of this Nation, produce large numbers of high performance students. However, given the nature of the social milieu from which these students typically come, the level of education of their parents, the efforts frequently spent in their homes to prepare them for school, and the many cultural and educational advantages they have by virtue of their community setting, it would be surprising indeed if such high expenditure schools did not produce highly capable students.

In time the above-described weaknesses of the cost-quality type of research became evident, and a new line of inquiry began. This time, the primary actors were those trained in methods of sociological research. The findings of these researchers, best illustrated perhaps in studies conducted by Alan B. Wilson and James S. Coleman,[4] tend to emphasize the significance of the student's social context, rather than school services, as determinants of pupil performance.

The general tenor of such sociological studies has been to demonstrate that a student's achievement is tied very tightly to his socioeconomic status. For example, in *Equality of Educational Opportunity,* differences were reported between ethnic groups as to their "sensitivity" to the effects of school quality.[5] On balance, however, in the view of Coleman and his fellow authors, the school service variables succeeded in explaining such a small portion of the variation in pupils' performance that they were moved to write:

> Taking all these results together, one implication stands out above all: That schools bring little influence to bear upon a child's achievement that is independent of his background and general social context; and that this very lack of independent effect means that the inequalities imposed upon children by their home, neighborhood, and peer environment are carried along to become the inequalities with which they confront adult life at the end of school.[6]

Critics of the Coleman Report hold that this conclusion is not necessarily warranted.[7] Their criticisms are at three levels: (1) inadequacy of the measurements utilized, (2) imprecise manipulation of those measures, and (3) inappropriate statistical techniques. Criticism one is exemplified by the Report's measures of school facilities, volumes-per-student in the school library and (for grades 9-12) the presence or absence of science laboratories. The critics' contention is that so few and such simple measures are insufficient in any attempt to understand the significance of the school in explaining pupil performance.

Criticism number two is exemplified by the treatment accorded the statistic "instructional expenditures per pupil." Each student was assumed by the Report to be benefiting from an annual instructional expenditure equal to the mean for his school district. The use of such an average masks intradistrict disparities, and from evidence displayed elsewhere in the Report such disparities appear to be substantial. By averaging expenditures and curtailing their distribution, the Report weighted the data against the possibility of finding a significant relationship.

The third major criticism involves the Report's statistical analyses. The issue here is that the Report's authors employed a form of regression analysis which is inappropriate if there exists a high degree of intercorrelation among "independent" variables. The Coleman Report attempted to explain variance in achievement scores by adding successively different independent variables to the analysis. The outcomes of this approach are highly sensitive to the order in which the explanatory variables are entered whenever the explanatory variables are interrelated.

The critics argue that Report measures of socioeconomic conditions and school services are highly interrelated and do not meet the criterion of independence. The argument here is that high quality school services tend to be made available to students from higher socioeconomic strata and lower quality school services to students from low socioeconomic strata.[8] If in a regression analysis "independent" variables are in fact highly intercorrelated, whichever variable cluster (socioeconomic status or school services) is first placed in the equation will have the highest explanatory power. The first entered cluster will have exhausted the major portion of whatever variance exists to be explained by the total of the two variable clusters together. The analysis involved in the Coleman Report chose to place socioeconomic status variables into the equation first; not unexpectedly they "discovered" that this cluster explained substantially more variance than did the school service cluster. Had they reversed the entry position of the two clusters, they would have found schools to be the major contributor to pupil performance.[9]

Studies which have emphasized, or overemphasized, the influence of social environment at the expense of school services, if taken on their face, have the effect of discounting the significance of schooling. At the other extreme, the cost-quality type study has frequently been oversimplified and construed to mean that schools will solve the problems of low pupil performance if only we spend more money. Clearly, in order to assess the determinants of intellectual achievement, or any other kind of student performance, adequate account must be taken of both the social context enveloping the student and the character of the school services to which he is exposed. Ideally, such an assessment should be of a "value added" nature. That is, we should like to determine what the child "knew" before he came to school, what he "knew" when he completed school, and how much of the difference was the unique contribution of the school. In order to conduct such an ideal study, the researcher would need to control methodologically for the possible influence of a host of out-of-school factors such as the student's innate intellectual capacity, family and home background, and neighborhood environment. Obviously, such total experimentation is presently impossible. Nevertheless, in this paper, we review research studies in which insofar as possible an attempt has been made to avoid the failings of past research in an effort to come closer to the "true" effects of schools upon student performance.[10]

A Perspective on Schooling

Before launching into research findings regarding the effects of various school services upon measures of pupil achievement, it seems appropriate to step back for a moment and attempt to gain a reasoned view of what it is that schools do and what it is that affects what schools do. Nowhere is it defined with precision, but schools in American society are expected to transform pupils on a large number of dimensions. A wide variety of attitudes, skills, and knowledge are expected to be "packed" into each pupil as a consequence of going to school. We do not yet understand well what mechanisms inside the human body enable one to "learn" these things. We do know, however, that whatever the process, or processes, they are extraordinarily complex. We can see this when we witness the wide range of ways in which children typically respond to the same events and stimuli. Children comprehend and express that comprehension in different ways, at different rates, and to varying degrees.

Whatever schools do to enhance this comprehension depends in a very major way upon the student's ability to perceive, store, process, and respond to a wide variety of environmental inputs. We do not, at least at this point, wish to become embroiled in what appears to be a specious argument as to whether this cluster of abilities is more sensitive to biological or environmental influences.[11] Suffice it here to say simply that almost all of the typical individual's biologically inherited components and a very substantial share of those which are environmentally shaped have taken hold prior to his first experiences with any formal education. Now, once having acknowledged the potential influence of genes and out-of-school environment, it seems reasonable to assume that the scope of variation in human performance which remains for the school to affect uniquely is somewhat limited. Moreover, it must be remembered that schools do not occupy the entire span of even the most ardent student's time. Even on a school day, and these frequently take up less than one-half of all the days in a year, a student is likely to be in the company and under the influence of his peers and parents for a longer period of time than he is engaged in school activities. Nevertheless, it still seems reasonable to expect the schools to have an effect; indeed, we will soon describe some of these effects.

But What Part of "School" Makes a Difference?

The term "school" is a deceptive generic label. *Webster's New World Dictionary* contains no less than 10 different contemporary definitions.[12] An etymological approach scarcely provides more precision. At its Latin roots, "school" refers to leisure, or the manner and location in which leisure took place. The difficulty with this ambiguity is that it complicates our desire to assess the "difference" that "school" makes. Only the most naive could possibly believe that the sheer act of being physically present in some building labeled SCHOOL renders an individual knowledgeable or skilled. Presumably, some sort of pedagogical process must be undergone before educational objectives are met. But just what are these processes? Where is it in the little "black box" labeled school that we should look? Is it the edifice itself? Is it the blackboards, the teacher, the textbooks, the movie projector, or the principal? Is it all of these things, or is it something else again?

In this quest, we are reminded of the frequent admonition: "Get the facts!" All right, but what facts? Facts about what? What "facts" are relevant? Without some systematic theoretical guidance, the researcher must resort to an almost random inquiry to isolate the essential ingredients. The plight is not quite this bad, we are able to resort to logic and prior research findings in order to identify school service components worthy of being tested for effectiveness. Nevertheless, the quest would be greater aided if we had a body of theory, theory about learning and instruction, which could guide us. Psychologists are daily discovering more about the nature of the learning process. We are perhaps still a long way from a unified theory of learning, but bits and pieces of such a theory are beginning to fall into place. What is not yet evolving very rapidly is a theory of instruction.[13] An analogy with the practice of medicine may be helpful in understanding the difference. To have a theory or body of knowledge which explains the origin of some particular disease is crucial to, but by itself insufficient for, treating a patient with that disease. Given knowledge that the patient has cancer, do you treat the illness with drugs, surgery, or radiation? This answer, of course, must rest upon the traits of the individual patient, the location and type of the cancer, the therapeutic processes at hand, and the skill of the physician. Much the same relationship holds between a learning theory which explains the processes which underlie reading and a teaching theory which would explain how to manipulate the environment to take advantage of the processes which "cause" one to be able to read. We are beginning to know moderately well the neurological and psychological mechanisms which interact to enable one to read. What we are just beginning to investigate is the means by which we can intervene in and manipulate those processes in the instance of individuals to make readers out of them. Given the biological and environmentally induced differences between individuals, the "treatment" for reading disabilities may well turn out to be complicated several-fold over the techniques necessary to treat cancer.

In the absence of a theory of instruction, educational researchers have typically tended to construct typologies of logically ordered school service components and to use available empirical measures to represent each of the typology categories. This is the general procedure followed in the research we will review. We do not wish to apologize for this nontheoretical approach or to bemoan *ad nauseum* the lack of an instructional theory. The point here is simply that research strategies based on "raw" empiricism are comparatively inefficient, and the continued lack of an instructional theory will hamper efforts to identify the *sine qua non,* the crucial instructional components, of schools.

Inability to construct a unified theory of instruction, however, has not been the only factor deterring identification of effective school service components. Another significant inhibitor of this quest has been the relatively slow development of research strategies and measurement methodologies applicable to education. Measures of output tend to be narrow; that is, they typically consist of a single performance criterion, for example, students' scores on various kinds of standardized achievement tests. Moreover, information about inputs is also frequently limited. The limitation here is that only a very few school systems collect information on any sizable number of significant input dimensions; and, even where such an effort is made, interdistrict comparisons are frequently frustrated by the lack of standardization in the data collected. Despite such handicaps, an increasing body of sophisticated research is accumulating on the effectiveness of various school service components, and we begin our review of such studies at this point. However, the reader who desires only a summary of this information can move directly to page 45 where we present a condensed version of these findings.

Research Findings

One of the forerunners in educational input-output analysis is a little known, but nevertheless significant, study done in 1956 for the Educational Testing Service by William G. Mollenkopf and S. Donald Melville.[14] These researchers gathered aptitude and achievement test scores from a nationwide sample of 9,000 ninth grade students in 100 schools and 8,357 12th grade students in 106 schools. Principals in each school responded to a questionnaire which led to the construction of 34 variables dealing with socioeconomic characteristics of students and their parents, availability of community provided educational opportunities, and quality of available school services. Given these three clusters of variables, the authors were able to assess the school's contribution to student performance while attempting to control for out-of-school influences. The authors are particularly careful to caution readers of the difficulty in prohibiting student socioeconomic status (SES) factors from contaminating any analysis of school service effects. Nevertheless, after controlling as best they could for student SES, they report four school service measures to be significantly related to pupil achievement. These are (1) number of special staff (psychologists, reading specialists, counselors, etc.) in the school, (2) class size, (3) pupil-teacher ratio,[15] and (4) instructional expenditures per student.

All of these findings suggest the central importance of the school staff and of students having relatively frequent contact with that staff. Measure number four is somewhat difficult to interpret because instructional expenditures usually include funds for supplies and equipment as well as staff salaries. However, in that the overwhelming proportion of this expenditure category is typically spent on instructional salaries, this measure also hints of the significance of the school's personnel in the learning of students. What is necessary now is to compare the results obtained in this study with those obtained in investigations where the controls for out-of-school influences are more adequate.

Another one of the early studies in this field was conducted in 1959 by the New York State Department of Education under the direction of Samuel M. Goodman.[16] This study, known as the *Quality Measurement Project,* covered a sample of 70,000 seventh and 11th grade students in 102 school districts selected for their ability to represent all of New York State. Findings here are comparable on two dimensions with the work of Mollenkopf-Melville. After partialing out the variance accounted for by the socioeconomic status of parents, Goodman reports per pupil instructional expenditures and number of special staff per 1,000 students to be significantly correlated with the achievement test scores of seventh grade students. In addition, two other characteristics were found to be significantly linked to pupil performance; they are teachers' experiences and a variable described as "classroom atmosphere." Teacher experience was measured as number of teachers in a district with 5 or more years of employment as a classroom instructor. "Classroom atmosphere" was a measure resulting from an observer's rating of the degree to which the teacher attempted to relate the subject matter under consideration to the interests and ability levels of students. In essence, it appears to be a measure of the degree to which the teacher was student oriented as contrasted with what educators frequently term "subject matter oriented." In general, Goodman's findings again point to the importance of the school's personnel in the instructional process.

J. Alan Thomas, in 1962, utilized Project TALENT information to test the impact of a large number of home, community, and school service variables upon student performance.[17] His sample was composed of 206 high schools in communities of 2,500 to 25,000 in 46 States. For 10th and 12th grade students in these schools he had scores on 18 separate achievement tests. Data about students, communities, and schools were taken from Project TALENT surveys and the 1960 census. Regression analysis was the statistical treatment utilized, and three measures of school service were taken to be significantly related with students' test scores, after taking home and community factors into account. These school service components are: (1) beginning teachers' salaries, (2) teachers' experience, and (3) number of volumes in the school library.

A unique examination of school effectiveness took place in 1964. It is not within the same analytical stream as the other studies we present, but it nevertheless warrants description. In the spring of 1959 the Board of Education in Prince Edward County, Va., voted to close all public schools under its authority. This action was taken in an effort to avoid the Supreme Court's racial desegregation decree. Thereafter, most white students in the County attended a segregated private school. Negro children, and a few poor whites, had several options: attend school in another county, participate in an assortment of volunteer efforts and makeshift schools, or forego formal education altogether. An inadvertent outcome of the school board's racist decision was to create some of the conditions necessary for an experimental analysis of school effectiveness. A team of Michigan State University researchers directed by Robert L. Green seized the opportunity.[18]

Significant differences were found in the home background and socioeconomic status of those children who attended schools *outside* the county. Thus they were excluded from comparison. However, no such out-of-school differences were found for those children who did and who did not participate in the *within* county volunteer schools. Participants and nonparticipants were administered standardized tests (Metropolitan Readiness and Stanford Achievement). Mean test scores were higher in almost every age group for those students who had participated in the intensive, formal, volunteer schooling programs. However, test score increments for age groups 6 to 10, though statistically significant, were minimal. For age groups 11 to 17, the gains were statistically significant and substantial.

A difficulty which arises in attempting to interpret this research is that the character of the educational services under study is imprecisely described and measured. Only the most gross kind of statement can be made: "Those children who attended the intensive volunteer educational program scored higher than those who did not." We do not know the nature of the educational program, and to that extent we are hampered in discovering the dimensions of schooling which account for learning.

Two significant studies of the effects of schools were reported in 1965: one, centered on schools in New York, was done by Herbert J. Kiesling[19] and the other, centered on schools in California was done for the California State Senate by Charles S. Benson.[20] The Benson study utilized data on fifth grade students from 249 school districts. Student performance was measured by standardized reading and mathematics tests. Data were compiled from the 1960 census on 12 socioeconomic and demographic variables of school district residents. Information was gathered from school districts and official statewide reports on 18 variables relating to school finance and expenditure allocations for school services. Because of a lack of time and the condition of the data, the study utilized only entire school districts, not individual schools, as the unit of analysis. Consequently, because of the averaging which occurs when measures for an entire district are used, the findings contain the potential to understate the importance of school service variables. Nevertheless, stepwise multiple regression analysis revealed teachers' salaries and instructional expenditures per pupil to be positively related to pupils' achievement even when socioeconomic status variables were taken into account. In Benson's words:

> The association between the achievement of pupils and the instruction offered by these teachers who are

> qualified by experience and training to be paid in the upper salary quartile is positive, and the association stands independently of the known connection between the home environment of pupils and their achievement.[21]

For medium-sized school districts (those with enrollments of 2,000 to 4,500 pupils) Benson found that, in addition to variables relating to teachers' salaries, mean salary of administrators was also positively associated with student achievement. Thus, from yet another study, we have strong evidence to suggest the importance of staff members with certain characteristics in influencing the performance of pupils.

The study of Kiesling utilized information collected in the previously described New York State Quality Measurement Project conducted by Goodman. One of Kiesling's major findings is that expenditures per pupil are positively related to student achivement (measured on Iowa Tests of Basic Skills and Iowa Tests of Educational Development). This finding holds specifically for large school districts (those with enrollments in excess of 2,000 pupils), particularly large urban school districts containing relatively large proportions of disadvantaged students. For For small districts, particularly small rural districts, the relationship between these two factors was frequently found to be random, and in some instances even to be negative. However, as the author is careful to suggest, the opportunity for various kinds of measurement idiosyncracies to manifest themselves is substantially greater in small districts. In a research sample composed of school districts which contain small numbers of students and very few teachers, the characteristics of individuals at the extremes of the measurement scales take on statistical significance out of proportion to their number. Moreover, as was the case with the Benson study, the per pupil expenditure variable used by Kiesling was a districtwide average figure and thus contains the potential to distort significantly the amount of resources spent on any individual student within a specific district. Nevertheless, one of the study's findings deserves particular emphasis. In Kiesling's words:

> The relationship of expenditure to performance in large urban districts is quite strong, with an additional $100 of expenditure being associated with 2.6 months of [achievement] at the beginning of the expenditure range and 1.4 months at the end of the range.[22]

In that the total per pupil expenditure figure for a school district represents money spent for a wide range of products and services, it is impossible to state precisely from Kiesling's findings just what school service component or components are making the difference. One extrapolation which appears reasonable, however, stems from the fact that the overwhelming portion of most school district's expenditures are for the salary of professional staff. (This figure typically accounts for from 65 to 85 percent of a school district's budget.) Consequently, it might be that the higher expenditure figure represents an ability to purchase services of instructional personnel who are more effective by virtue of their experience, preparation, and general ability. These increments in the quality of staff, in turn, reflect themselves in the achievement test scores of students. This is but a supposition, however, because Kiesling does not present data directly related to teacher preparation and experience.

Results of the study *Equality of Educational Opportunity* (the Coleman Report) were made public in 1966. At the beginning of this paper we noted the limitations of the Coleman team's efforts. At this point it is appropriate also to acknowledge some of the Report's strengths. The Coleman Report represents the most extensive attempt at assessment of a Nation's entire educational system ever made. The survey collected information on approximately 660,000 students attending thousands of schools in hundreds of school districts in every region of the United States. In addition, data were gathered regarding the teachers of those students, the characteristics of their schools, the range and diversity of their curriculums, qualifications of the school administrators, and so on. Because of serious measurement errors and inappropriate analytical procedures, we believe that Coleman and his colleagues, though unintentionally, underestimate the potential significance for pupil achievement of a number of the school service components they examined. Nevertheless, a fact which is worthy of emphasis is that, even having biased their analysis against finding effective school service components, the Coleman team does report several such components to be positively and significantly associated with pupils' performance.[23]

The most significant school service variable in explaining student achievement (measured by a vocabulary test) was a teacher characteristic, the teacher's verbal ability. As with the other findings of this nature that we have discussed, care must be used in interpreting the meaning of such a result. What the Coleman team reports is that, after having made an effort to control statistically for a student's home background and community social environment, his achievement test results tend to increase in relation to the verbal ability level of his teacher. Obviously, much more is involved in the instruction of a student than his teacher's skill at responding to verbal ability test questions. However, if one views teachers' verbal ability as a proxy measure for a number of related skills and qualities, the Coleman Report finding can be interpreted in a meaningful fashion.[24] If the measure of verbal ability is taken to represent the general intelligence level of the teacher, the finding can be construed to mean that an intellectually facile instructor is more adept at tasks such as finding means to motivate students, adapting materials to their ability levels, and communicating in ways which make the subject matter more understandable. This is an interpretation which is totally consistent with the observations and conventional wisdom of untold thousands who have themselves been teachers or who have supervised teachers.

An interesting adjunct to the Coleman finding about teachers' verbal ability is that the variable appears to have an accumulative effect. It is statistically significant when examined for sixth grade students and thereafter increases in importance when examined for ninth and 12th grade students. Moreover, its effect tends to vary in accord with the characteristics of the student. It shows consistently positive correlations with the achievement of all students, but it appears to be especially important in explaining the achievement levels of Negro students. To paraphrase the Coleman Report, Negro children appear to respond in a particularly sensitive and positive fashion to a teacher who is skilled verbally.

In the year following issuance of the Coleman Report (1967), three additional studies were published which deal with some facet of the topic of school service effectiveness. Two of these, a study by Marion F. Shaycoft[25] and a study directed by Jesse Burkhead[26] focus on U.S. secondary schools. The third study, the so-called Plowden Report,[27] was conducted in England.

The Shaycoft study is unusually informative on several dimensions and somewhat disappointing on some others. Its greatest asset results from the procedures employed to measure student performance. The study sample consisted of 6,583 students who were tested by Project TALENT in 1960 when they were in the ninth grade. Subsequently, these students matriculated to 118 different secondary schools (101 of which were comprehensive high schools, the other 17 were specialized vocational high schools).[28] In 1963 this same cohort of students was administered a battery of examinations designed for 12th grade students. The test battery, in addition to having the usual generalized tests of verbal and quantitative reasoning ability, also included achievement examinations in specific subject areas, e.g., foreign language, English, accounting, and literature. Presumably, schools are established to instruct students in moderately well-defined subject matter areas, not to increase some quality as amorphous as "verbal ability." Consequently, the Shaycoft output measures appear to be more related than those of most studies to the unique functions and objectives of schools.

A second favorable feature of the Shaycoft study is the use of longitudinal or time series testing. What a student knew about a particular subject was measured in grade nine, and this information was used as a baseline against which to assess increments in achievement for the subsequent 3 years of schooling. This procedure, more closely than most other methods, enables the researcher to gain a picture of the "value added" to the student during the course of his schooling. Moreover, in that the tests were heavily concentrated on school-related subjects, subjects about which one typically does not learn outside of schools, the room for alternative explanations of achievement gains is reduced.

The Shaycoft analyses reveal student achievement gains over the 3 years to be consistent and of a healthy magnitude. In most instances, 12th grade achievement gains represented a difference of one standard deviation when compared to ninth grade norms. This is so even when differences in students' socioeconomic status are controlled statistically. It is reasonable to infer from such a finding that for the schools in question some school service characteristics are influencing student achievement. The difficulty, and consequently the disappointment, with the Shaycoft study, is that only a very limited spectrum of school service components was examined. The study concentrated on the availability within schools of particular subject matter offerings. No measures of components such as staff quality, instructional material availability, or equipment and facility adequacy were employed. What can be said is that the availability of a particular curriculum in a school is related significantly to whether or not students grew in knowledge about the subject matter contained in that curriculum. Not surprisingly, for example, when schools did not offer courses in accounting or electricity, then students' scores on achievement tests in these areas were limited.

The effort by Burkhead and his colleagues lacks the richness of the Shaycoft study on the dimension of subject matter output measures, but it is much more complete in terms of the school service components it examines. The Burkhead study sample included 39 Chicago public secondary schools (enrolling almost 90,000 students), and 22 Atlanta public high schools (enrolling a total of approximately 19,000 students). Results for schools in these two large cities were compared with data from a Project TALENT sample of approximately 180 public high schools in smaller communities. Information regarding students' performance was constructed from scores on a variety of tests of aptitude, reading, and general knowledge, and measures of school persistence (the degree to which students do *not* "drop out" of schools). Socioeconomic status measures were derived from 1960 census data about residents in high school attendance areas. School service components consisted of measures such as teacher-man-years per pupil, teachers' experience, and school building age. Statistical techniques were employed in an effort to control for the SES of students. Unfortunately, however, these statistical procedures were essentially the same as those employed by the Coleman Report team, and, thus, tend to understate seriously the potential impact of school service components. Nevertheless, as with the Coleman Report, despite methodological limitations biasing the findings against schools, Burkhead reports some school services to be effective.

Findings varied somewhat from Chicago to Atlanta, probably, at least in part, reflecting the lack of standardization in the input and output measures available for schools in the two cities. Moreover, results from analyses of Chicago's schools were somewhat hampered by lack of variation or dispersion in the quality of school services dispensed at the different schools. Nevertheless, in Chicago, newer buildings were found to be associated with lower dropout rates and the teacher's experience was linked to pupils' reading scores. For Atlanta schools, low rates of teacher turnover were found to be positively associated with increments in pupils' scores on tests of verbal ability. For the sample of high schools in small communities, the beginning salary and years of experience for teachers and the age of the school building were found to explain variations in test score results.

The previously referred to work of England's Central Advisory Council on Education (The Plowden Report) consists of two volumes, volume I presents the policy recommendations of the Council and volume II contains results of the several research studies which serve to support these recommendations.[29] For our purposes, the Plowden Report's most significant research study is the *National Survey of Parental Attitudes and Circumstances Related to School and Pupil Characteristics,* directed by Gerald Peaker. This effort collected information from a stratified random sample of primary school students as to academic performance and school and home characteristics. These data enabled the study team to assess the relative influence upon pupil performance of home and socioeconomic status characteristics and school service components. The primary statistical procedure employed was regression analysis.

Except for the fact that the study limits itself to a concern for elementary school students, its findings and the controversies surrounding them are not very different from those which have accompanied the Coleman Report in this Nation. Nevertheless, several school service components are described as contributing in a statistically significant fashion to an explanation of pupil achievement. These components deal with the school building and the teacher. Specifically, age of building and teacher's experience, academic preparation, and "ability" were found to be positively associated with output measures. These findings are all consistent with and support the results of the several other studies we have already reviewed.

Added evidence of the significant role played by teachers in the instructional process is provided in a 1968 study by Elchanan Cohn.[30] As an economist, Cohn was primarily concerned with examining possible economies of scale in public high school operations. His analyses, however, also lend themselves to our search for information about the effectiveness of various school service components. For secondary school students in a sample of 377 school districts in the State of Iowa, Cohn obtained information relative to achievement (as measured by scores on the Iowa Test of Educational Development) and school services (mostly expenditure data and information about reacher characteristics). Using multiple regression analysis, Cohn reports that amount of teacher salary and number of instructional assignments per teacher are associated with increments of pupil achievement, and the direction of the association is in keeping with conventional expectations. The higher the salary and the fewer the number of different reaching assignments for a teacher, the higher the test scores of pupils. In terms of his primary objective, assessing economies of scale, Cohn found high schools with enrollments between approximately 1,250 and 1,650 students to be the most cost-effective.

The extent to which Cohn's study utilized an effective statistical control for certain nonschool inputs (student aptitude and SES) is questionable. Consequently, the results in terms of the unique contribution of school services must be interpreted with caution. Nevertheless, Cohn's findings are consistent with what we have come to expect by comparison with findings from other studies.

A study somewhat similar to Cohn's was reported in 1968 by Richard Raymond.[31] Raymond's sample consisted of approximately 5,000 West Virginia high school students who graduated between 1963 and 1966 and who subsequently matriculated to the University of West Virginia. The freshman year performance of these students was measured by achievement test scores and individual grade point averages. Students were grouped by the county in which their high school was located, and measures of school service characteristics were then obtained for county school systems.[32] Four measures of socioeconomic status for the residents of these counties were obtained from 1960 census data. Using these census figures to control for SES, Raymond regressed school service components on the two output measures and found teachers' salaries to explain a significant portion of the variance in students' freshman year scholastic performance. The salaries of elementary school teachers appeared to be particularly powerful variables in explaining differences in student achievement.

A portion of the 1968 study of Boston schools done by Martin Katzman examines the relationship between school services and student achievement.[33] He collected data from 56 of the Boston school system's elementary school attendance districts. Information was gathered on six dimensions of pupil performance: three measures having to do with regularity of attendance and school holding power and three scholastic measures (percentage of students taking and percentage passing the entrance examination to the city's academically elite Latin High School, and reading achievement increments as determined by differentials between second and sixth grade examination results).

Using multiple regression analysis in an effort to control for students' SES, Katzman found school service variables to be significantly associated with one or more of the above output measures in the following fashion:

A measure of "crowding" was derived from the number of classrooms which contained more than 35 students. That figure represented the modal number of desks in Boston city schools' classrooms; students in excess of this number were taken to be in some sort of makeshift arrangement. The consequences of crowding were not found to be clear and consistent on the attendance output measures. Noncrowding, however, was associated with increments of reading achievement and number of students passing the Latin High School's entrance examination.

The ratio of students to staff members was found to have consistent and significant correlation with school attendance and school persistence output measures.

The size of the attendance district appeared to provide some economies of scale when judged on the output criteria of reading scores and school persistence. That is, the larger the number of children served by an attendance district, the higher their reading achievement increments and the greater the schools' holding power. However, in contrast to these positive consequence of size, some diseconomies of scale were found when the output measures dealt with the Latin High School. The larger the attendance district's enrollments, the smaller the proportion of students who sat for and passed the Latin High School's entrance examination.

The percentage of permanently employed teachers was found to have minor, but nevertheless positive, effects on all output measures. The greater the percentage of permanently employed teachers, tenured teachers, the better the performance of pupils.

Percent of teachers who possessed a master's degree was found to have generally positive effects. This component demonstrated particularly strong relationships with measures of school attendance.

The percent of teachers in an attendance district with from one to 10 years of teaching experience was taken as a school service component or input variable. The relationship of this measure to outputs was interesting, but inconsistent. Experience was positively associated with measures of school attendance and holding power, but negatively related to relative increments in reading achievement.

The turnover rate among teachers within an attendance district was demonstrated to have a slight negative association with all the output measures.

Katzman's study adds substantially to the evidence supporting the significant role of school staff in effecting pupil performance. As with almost all such efforts, however, the findings of his study would be even more helpful had he been able to enlarge the scope and refine the input measures considered. The finding for teacher experience provides an interesting example here. To know that the variable "percent of teachers with from 1 to 10 years of teaching experience" is positively linked to increments in holding power, but negatively associated with relative increments of reading achievement is to paint a somewhat perplexing picture. If Katzman had had access to detailed information, we could begin to see more precisely whether these findings result from very new, inexperienced teachers, say in their first year, or teachers near the 9- and 10-year end of the category.

In 1968, Samuel Bowles presented preliminary results of another study on educational production functions.[34] Bowles' findings are based on a sample of 12th grade Negro male students constructed from data compiled by the Coleman Report survey team. Bowles is careful to circumscribe the validity and generalizability of his findings by referring to the limitations of the sampling and measurement procedures employed in the initial collection of the data. Despite these limitations, we find his results to be of interest; not only do they reaffirm the significance of teacher characteristics, but also they suggest certain additional categories of school service components to be important. Regression analysis was employed, and four measures of a student's home environment were entered into the equation in an effort to control for out-of-school influences. The relative presence of science laboratory facilities, the average amount of time a teacher spends in guidance activities, and the number of days the school stays in session during a school year are all variables found to be significantly associated with students' scores on tests of verbal ability. The "science teaching laboratory" variable is somewhat similar to "teacher's verbal score" in that it needs to be interpreted. How can the presence or absence of science laboratories have an impact on student achievement when the latter is measured by general tests of reading and vocabulary? Our answer to this query is to take science laboratories as a proxy measure of school facilities more generally. The logic here is that schools possessing such laboratory facilities are also likely to be relatively well supplied on most other dimensions of school facilities. Conversely, a school lacking science laboratories is also likely to be in a poor position with regard to other facilities used for instruction.

In another place, Bowles reports findings from a study which utilized a sample of 12th grade Negro students for which Project TALENT information was available.[35] In this instance, the output measures were students' achievement in mathematics and reading and scores on a test of generalized academic ability. Bowles found large class size and "teaching" or ability grouping to be negatively related and amount of teachers' graduate preparation to be positively related to students' performance on reading tests. Only the class size variable was significant at the .05 level, however. When mathematics achievement scores were taken as the criterion, ability grouping and age of school building appeared to have a negative influence and expenditures per pupil and teachers' graduate preparation a positive influence. Finally, on the test of general academic ability, class size and ability grouping were again found to be negatively related and teacher preparation level positive. All of these findings came about after statistical controls for students' social environment had been exercised.

In another study, coauthored with Henry Levin, Bowles presents more findings about the effectiveness of several other school service components.[36] During the course of their literary debate with James S. Coleman and his colleagues regarding the validity of findings presented in *Equality of Educational Opportunity,* Bowles and Levin employed EEO data in a regression analysis which attempted to correct for some of the Coleman

Report's controversial methodological procedures. These analyses were conducted using verbal ability test results as output measures for 12th grade Negro students. In this effort, they found teachers' salaries and science laboratories to be significantly related to pupil performance. In another regression analysis in the same study, they found teachers' verbal ability to be significantly related to student achievement. These same findings held generally for analyses done for white 12th grade students, but, for reasons which are not readily explainable, the levels of significance were lower.

Somewhat similar findings stem from a 1968 study done by Eric Hanushek.[37] This study attempts to calculate educational production functions for sixth grade children using standardized achievement test scores as a criterion of output and measures derived from Coleman Report data as inputs. The study centers on white children in 471 elementary schools and Negro children in 242 elementary schools in the metropolitan North. Regression analysis was the statistical procedure utilized with suitable controls for socioeconomic status. Significant relationships to achievement were found for teachers' verbal ability and years of teaching experience.

Also in 1968, Thomas I. Ribich published the results of a study utilizing information from Project TALENT.[38] Ribich's procedure was to examine only those students who fell into the lowest quintile on measures of socioeconomic status. When this control was exercised for out-of-school influences, it was found that pupils' performance on standardized achievement tests was directly related to expenditures per pupil.

In 1969, Guthrie and his colleagues conducted an assessment of school effectiveness using data collected in Michigan for the Equal Educational Opportunity Survey.[39] In an effort to avoid the methodological problems previously described for the Coleman Report findings on school effectiveness, a different analytical technique was employed. The sample consisted of 5,284 sixth grade students, both Negro and white. A socioeconomic status score for each student was computed from information regarding parental income and education. These scores were hierarchically ordered and subsequently divided into 10 equal groups. Each decile subset contained approximately 528 students who were relatively homogeneous with regard to their social background. Separate analyses were then conducted for each decile in order to assess the relationship between measures of school service quality and student scores on tests of reading ability, mathematics understanding, and verbal facility

In these analyses, a total of 11 school service variables were found to relate significantly to students' performance measures. The school service variables are listed below by category.

School Facilities
  a. School site size
  b. Building age
  c. Percent of makeshift classrooms

Instructional Materials
  a. Library volumes per student
  b. Supply of textbooks

Teacher Characteristics
  a. Verbal ability
  b. Experience
  c. Job satisfaction

Student Environment
  a. School size (enrollment)
  b. Classrooms per 1,000 students
  c. Percent of students transferring

Summary of Effective School Service Components

In the preceding section we reviewed 19 studies which deal with the effectiveness of school service components. These investigations have been conducted using a variety of sample subjects, input and output measures, and controls for what are commonly presumed to be out-of-school influences upon pupil performance. In order to impose some degree of uniformity upon this diversity, we have attempted to condense the essential components of each investigation into a summary chart which appears at the end of this chapter.

From an inspection of these digested results it is evident that there is a substantial degree of consistency in the studies' findings. The strongest findings by far are those which relate to the number and quality of the professional staff, particularly teachers. Fifteen of the studies we review find teacher characteristics, such as verbal ability, amount of experience, salary level, amount and type of academic preparation, degree level, job satisfaction, and employment status (tenured or nontenured), to be significantly associated with one or more measures of pupil performance.

In order for school staff to have an effect upon students, however, it is necessary that students have physical access to such persons. And, indeed, we also find that student performance is related to some degree to contact frequency with or proximity to professional staff. This factor expresses itself in variables such as student-staff ratios, classroom size, school or school district size, and length of school year.

In addition to findings in support of the effectiveness of staff, a number of studies under review also present results to suggest that service components such as age of school building, adequacy and extent of physical facilities for instruction also are significantly linked to increments in scales of pupil performance. Finally, as might be expected logically because all the foregoing components translate into dollar costs, we find that measures such as expenditures per pupil and teachers' salary levels correlate significantly with pupil achievement measures.[40]

Conclusion

In conclusion, we are impressed with the amount and consistency of evidence supporting the effectiveness of school services in influencing the academic performance of pupils. In time, we would wish for more precise information about which school service components are most effective and in what mix or proportion they can be made more effective. Nevertheless, on the basis of information obtained in the studies we review, there can be little doubt that schools do make a difference.

Summary Chart of Effectiveness Studies on School Service Components

Study Author(s)	Description of Sample	Measure of Pupil Performance (School Output)	Measure(s) of Effective School Service Component(s) (School Input)
1. Mollenkopf and Melville	U.S., 17,000 9th (in 100 schools) and 12th (in 106 schools) grade, male and female	Aptitude and achievement tests	1. Number of special staff 2. Class size 3. Pupil-teacher ratio 4. Instructional expenditures
2. Goodman	New York, 70,000 7th and 11th grade, male and female in 102 school districts	Achievement test	1. Number of special staff 2. Instructional expenditures 3. Teachers' experience 4. "Classroom atmosphere"
3. Thomas	Project TALENT Sample (national) 10th and 12th grade, male and female	Achievement test	1. Teachers' salaries 2. Teachers' experience 3. Number of library books
4. Green, *et al.*	Virginia (Primarily Negro) Secondary students	Stanford Achievement Test	1. Aggregate measure of entire instructional program
5. Benson	California 5th grade, 249 school districts	Reading achievement test	1. Teachers' salaries 2. Administrators' salaries 3. Instructional expenditures
6. Kiesling	New York, 70,000 7th and 11th grade male and female in 102 school districts	Achievement test	1. Expenditure per pupil (in large school districts)
7. Coleman Report	U.S. sample	Verbal ability test	1. Teachers' verbal ability
8. Shaycoft	U.S. 108 schools 6,500 9th and 12th grade, male and female	Battery of 42 aptitude and achievement tests	1. Curriculum variables
9. Burkhead	90,000 Chicago high schools students in 39 schools. 19,000 Atlanta High School students in 22 schools and 180 small community high schools	Aptitude and achievement tests and school holding power	1. Age of building 2. Teachers' experience 3. Teacher turnover 4. Teachers' salary
10. Plowden Report	English elementary school students		1. Age of building 2. Teachers' experience 3. Teachers' academic preparation 4. Teachers' "ability"
11. Cohn	Iowa high school students in 377 school districts	Achievement test	1. Teachers' salary 2. Number of instructional assignments per teacher 3. School size
12. Raymond	W. Virginia 5,000 high school students	Freshman year (college) GPA and achievement test scores	1. Teachers' salary

Summary Chart of Effectiveness Studies on School Service Components—*(Continued)*

Study Author(s)	Description of Sample	Measure of Pupil Performance (School Output)	Measure(s) of Effective School Service Component(s) (School Input)
13. Katzman	Boston elementary school students	School attendance, school holding power, Reading achievement, Special school entrance examination	1. Pupils per classroom 2. Student-staff ratio 3. Attendance district enrollment 4. Teachers' employment status 5. Teachers' degree level 6. Teachers' experience 7. Teacher turnover ratio
14. Bowles (1)	U.S. 12th grade Negro males	Verbal ability test	1. Teachers' verbal ability 2. Science laboratory facilities 3. Length of school year
15. Bowles (2)	U.S. 12th grade Negro males	Mathematics and reading achievement test and a test of general academic ability	1. Class size 2. Ability grouping 3. Level of teacher training 4. Age of school building 5. Expenditures per pupil
16. Bowles & Levin	12th grade Negro students and 12th grade white students	Verbal ability test scores	1. Teachers' verbal ability 2. Teachers' salary
17. Hanushek	6th grade white students in 471 schools and 6th grade Negro students in 242 schools	Verbal ability test	1. Teachers' verbal ability 2. Teachers' experience
18. Ribich	Project TALENT Sample	Achievement test	1. Expenditures per pupil
19. Guthrie, *et al.*	5,284 6th grade students in Michigan	Reading ability, Mathematics understanding, Verbal facility	1. School site size 2. Building age 3. % classrooms makeshift 4. Library volumes 5. Textbook supply 6. Teachers' verbal ability 7. Teachers' experience 8. Teachers' job satisfaction 9. School size (enrollment) 10. Classrooms per 1,000 students 11. % of students transferring

Acknowledgments

This paper is adapted from chapter 4 of *Schools and Inequality*, a study conducted in 1969 by the author together with George B. Kleindorfer, Henry M. Levin, and Robert Stout for the Urban Coalition.

Footnotes

[1] Coleman, James S., *et al., Equality of Educational Opportunity* (Washington, D.C.: U.S. Government Printing Office, 1966).

[2] A review of the cost-quality line of inquiry and some of its successors is provided by William E. Barron's chapter, "Measurement of Educational Productivity," in *The Theory and Practice of School Finance,* edited by Warren E. Gauerke and Jack R. Childress (Chicago: Rand McNally Co., 1967), pp. 279-308. An earlier review of such efforts is provided in Mort, Paul R., "Cost Quality Relationships in Education." *Problems and Issues in Public School Finances,* edited by R. L. Johns and Edgar L. Morphet (New York: National Conference of Professors of Educational Administration, 1952).

[3] See, for example, Furno, Orlando Frederick, "The Projection of School Quality from Expenditure Level" (unpublished doctoral dissertation, Columbia University, 1956).

[4] In this context, one can take, for example, either the Coleman Report to which we have already referred or an earlier study by the same author, "The Adolescent Subculture and Academic Achievement," *The American Journal of Sociology,* Volume 65 (1960), pp. 337-347. An excellent example of Wilson's research is "Residential Segregation of Social Classes and Aspirations of High School Boys," *American Sociological Review,* Volume 24 (1959), pp. 836-845.

[5] Negroes, Indian-Americans, Mexican-Americans, and Puerto Ricans tended to respond more dramatically to contact with good teachers and enriched programs than did white students.

[6] *Equality of Educational Opportunity,* p. 325.

[7] For a more detailed explanation of the limitations of Coleman Report findings, see Bowles, Samuel S. and Levin, Henry M., "The Determinants of Scholastic Achievement: An Appraisal of Some Recent Findings," *Journal of Human Resources,* Volume III, No. 1 (Winter 1968).. Also, see "More on Multicollinearity and the Effectiveness of Schools," *Journal of Human Resources,* Volume 3, No. 5 (Summer 1968), by the same authors.

[8] Strong evidence for this proposition is provided in the research reported in chapter 3 of *Schools and Inequality.*

[9] Bowles and Levin, *op. cit.*

[10] For further elaboration upon the difficulties inherent in assessing the effect of schools see Werts, Charles E. and Linn, Robert L., "Analyzing School Effects: How to Use the Same Data to Support Different Hypotheses," *American Educational Research Journal,* Volume VI, No. 3 (May 1969), pp. 439-447.

[11] See, for example, the article by Jensen, Arthur R., "How Much Can We Boost IQ and Scholastic Achievement?" *Harvard Educational Review,* Volume 39, No. 1 (Winter 1969), and the critical reactions to it in the subsequent issue, Volume 39, No. 2 (Spring 1969).

[12] *Webster's New World Dictionary of the American Language,* College Edition (New York: World Publishing Company, 1966), p. 1304.

[13] The need for a theory of instruction is forcefully explained in an article by Nathan L. Gage entitled "Theories of Teaching" in *Theories of Learning and Instruction*, The Sixty-Third Yearbook of the National Society for the Study of Education (Chicago: University of Chicago Press, 1964), pp. 268-285.

[14] Mollenkopf, William G. and Melville, S. Donald, "A Study of Secondary School Characteristics as Related to Test Scores," Research Bulletin 56-6 (Princeton: Educational Testing Service, 1956), mimeograph.

[15] The second and third measures (class size and pupil-teacher ratio) represent similar but not identical phenomena. For example, it is possible for a school to have a relatively high ratio of pupils to teachers, but if each teacher instructs in six or more classes, average class size may be relatively low. In general, however, where class size is large there will be relatively few staff members for the number of students enrolled.

[16] Goodman, Samuel M., *The Assessment of School Quality* (Albany: The State Education Department of New York, 1959).

[17] Thomas, J. Alan, "Efficiency in Education: A Study of the Relationship between Selected Inputs and Mean Test Scores in a Sample of Senior High Schools," unpublished PhD. dissertation (Stanford University: School of Education, 1962).

[18] Green, Robert Lee, *et al.*, "The Educational Status of Children in a District Without Public Schools," Bureau of Educational Research Services, College of Education, Michigan State University, U.S. Department of Health, Education and Welfare, Office of Education Cooperative Research Project No. 2321, 1964.

[19] Kiesling's study was an unpublished Harvard University Ph.D. dissertation. The results of that study are more readily available in an article entitled "Measuring a Local Government Service: A Study of School Districts in New York State," *Review of Economics and Statistics*, Volume XLIX, No. 3 (August 1967), pp. 356-367.

[20] Benson, Charles S., *et al.*, *State and Local Fiscal Relationships in Public Education in California*, a report of the Senate Fact Finding Committee on Revenue and Taxation published by the Senate of the State of California, March 1965.

[21] *Ibid.*, p. 56.

[22] Kiesling, *op. cit.*, p. 365. The word "achievement" in this quotation is ours. The journal article has the word "expenditure" at that exact point, but the meaningless nature of the term in that context leads us to believe that it is a printing error and that our substitution is consistent with the author's intent.

[23] The analyses of the effect of school service components upon pupil performance is discussed in the Coleman Report from page 290-332. In addition to the already cited works by Bowles and Levin, anyone who is deeply interested in studies of school effectiveness should read Cain, Glen and Watts, Harold, "Problems in Making Inferences from the Coleman Report," mimeographed working paper of the Institute for Research on Poverty (Madison: University of Wisconsin, 1968), and Kain, John F. and Hanushek, Eric A., "On the Value of Equality of Educational Opportunity as a Guide to Public Policy," mimeographed working paper #36 of the Program on Regional and Urban Economics (Cambridge: Harvard University, 1968).

[24] For additional information on the relationship of verbal ability to other personal attributes, see Flanagan, John C., *et al.*, *The American High School Student* (Pittsburgh: Project TALENT office, University of Pittsburgh, 1964), chapters 7 and 8.

[25] Shaycoft, Marion F., *The High School Years: Growth in Cognitive Skills* (Pittsburgh: American Institute for Research and School of Education, University of Pittsburgh, 1967).

[26] Burkhead, Jesse, Fox, Thomas G. and Holland, John W., *Input and Output in Large City High Schools* (Syracuse: Syracuse University Press, 1967).

[27] This study represents the efforts of a distinguished committee chaired by Lady Plowden. The research study and report were issued by the Central Advisory Council

on Education and are officially entitled *Children and Their Primary Schools* (London: Her Majesty's Stationery Office, 1967).

[28] The secondary schools were selected on the basis of a stratification procedure which aimed at constructing a sample which was representative of all secondary schools in the Nation.

[29] A discussion and critique of both volumes is provided in separate articles by Joseph Featherstone and David Cohen in the *Harvard Educational Review,* Volume 38, No. 2 (Spring 1968), pp. 317-340.

[30] Cohn, Elchanan, "Economies of Scale in Iowa High School Operations," *Journal of Human Resources,* Volume 3, No. 4 (Fall 1968), pp. 422-434.

[31] Raymond, Richard, "Determinants of the Quality of Primary and Secondary Public Education in West Virginia," *Journal of Human Resources,* Volume 3, No. 4 (Fall 1968), pp. 450-470.

[32] As with most Southern States, in West Virginia the county serves as the primary unit for organizing local school districts.

[33] Katzman, Theodore Martin, "Distribution and Production in a Big City Elementary School System," *Yale Economic Essays,* Volume 8, No. 1 (Spring 1968), pp. 201-256.

[34] Bowles, Samuel S., "Towards an Educational Production Function," mimeographed paper presented at the Conference on Research in Income and Wealth (November 1968). (To be published in a forthcoming volume entitled *Income and Education,* edited by W. Lee Hansen.)

[35] Bowles, Samuel S., "Educational Production Functions," Final Report to the Office of Education under cooperative research contract OEC 1-7-00451-2651 (February 1969), (see especially the tables on pp. 61-63).

[36] Bowles, Samuel S. and Levin, Henry M., "More on Multicollinearity and the Effectiveness of Schools," *Journal of Human Resources,* Volume 3, No. 3 (Summer 1968), pp. 393-400.

[37] Hanushek, Eric, "The Education of Negroes and Whites," unpublished doctoral dissertation (Department of Economics, Massachusetts Institute of Technology, 1968).

[38] Ribich, Thomas I., *Education and Poverty* (Washington, D. C.: Brookings Institution, 1968).

[39] Reported in *Schools and Inequality* and to be described in detail in a paper prepared for the American Educational Research Association Annual Meeting in Minneapolis, March 2-5, 1970.

[40] For a more detailed description of the manner in which teacher quality characteristics translate into dollar costs, see Levin, Henry M., *Recruiting Teachers for Large City Schools* (New York: Charles Merrill and Sons, 1970).

References

Barron, William E., "Measurement of Educational Productivity," *The Theory and Practice of School Finance,* edited by Gauerke, Warren E., and Childress, Jack R. (Chicago: Rand McNally Co., 1967).

Benson, Charles S., *et al., State and Local Fiscal Relationships in Public Education in California,* Report of the Senate Fact Finding Committee on Revenue and Taxation (Sacramento: Senate of the State of California, March 1965).

Bloom, Benjamin S., "International Project on the Evaluation of Educational Achievement," Bulletin No. 4, UNESCO Institute for Education (Hamburg, 1964).

Bowles, Samuel S., "Educational Production Functions," Final Report to the Office of Education under Cooperative Research Contract OEC 1-7-00451-2651 (February 1969).

Bowles, Samuel S., "Toward an Educational Production Function," paper presented at the Conference on Research in Income and Wealth, November 15-18, 1968 (to be published in a forthcoming volume entitled *Income and Education,* edited by Hansen, W. Lee).

Bowles, Samuel S., and Levin, Henry M., "More on Multicollinearity and the Effectiveness of Schools," *Journal of Human Resources,* Vol. 3, No. 3 (Summer 1968).

Bowles, Samuel S., and Levin, Henry M., "The Determinants of Scholastic Achievement: An Appraisal of Some Recent Findings," *Journal of Human Resources,* Vol. 3, No. 1 (Winter 1968).

Burkhead, Jesse, Fox, Thomas G., and Holland, John W., *Input and Output in Large City High Schools* (Syracuse: Syracuse University Press, 1967).

Cain, Glen, and Watts, Harold, "Problems in Making Inferences from the Coleman Report," mimeographed working paper of the Institute for Research on Poverty (Madison: University of Wisconsin, 1968).

Central Advisory Council for Education, *Children and their Primary Schools* (London: Her Majesty's Stationery Office, 1967).

Cohn, Elchanan, "Economies of Scale in Iowa High School Operations," *Journal of Human Resources,* Vol. 3, No. 4 (Fall 1968).

Coleman, James S., "The Adolescent Subculture and Academic Achievement," *The American Journal of Sociology,* Vol. 65 (1960).

Coleman, James S., *et al., Equality of Educational Opportunity* (Washington, D.C.: U.S. Government Printing Office, 1966).

Crandall, James Henry, "A Study of Academic Achievement and Expenditures for Instruction," (unpublished Ed.D. dissertation, University of California, Berkeley, 1961).

David, M., Brazer, H., Morgan, J., and Cohen, W., *Educational Achievement—Its Causes and Effect* (Ann Arbor: University of Michigan, 1961).

Furno, Orlando Frederick, "The Projection of School Quality from Expenditure Level" (unpublished doctoral dissertation, Columbia University, 1956).

Gintis, Herbert, "Production Functions in the Economics of Education and the Characteristics of Worker Productivity" (unpublished doctoral dissertation, Harvard University, 1969).

Goodman, Samuel M., *The Assessment of School Quality* (Albany: The State Education Department of New York, 1959).

Green, Robert Lee, *et al.*, "The Educational Status of Children in a District Without Public Schools," Bureau of Educational Research Services, College of Education, Michigan State University, U.S. Department of Health, Education, and Welfare, Office of Education Cooperative Research Project No. 2321, 1964.

Guthrie, J. W., *et al., Schools and Inequality* (Washington, D.C.: The Urban Coalition, 1969).

Guthrie, J. W., *et al.*, "A Study of School Effectiveness: A Methodological Middleground," paper to be presented to the American Educational Research Association, March 2-5, 1970.

Hanushek, Eric A., "The Education of Negroes and Whites" (unpublished doctoral dissertation, Department of Economics, Massachusetts Institute of Technology, 1968).

Hawkridge, D., Chalupsky, A., and Roberts, A.O.H., "A Study of Selected Exemplary Programs for the Education of Disadvantaged Children," Final Report, U.S. Office of Education Project No. 089013 by American Institute of Research, Contract No. OEC-0-8-089013-3515 (010), (September 1968).

Kain, John F., and Hanushek, Eric A., "On the Value of Equality of Educational Opportunity as a Guide to Public Policy," mimeographed working paper #36 of the Program on Regional and Urban Economics (Cambridge: Harvard University, 1968).

Katzman, Theodore Martin, "Distribution and Production in a Big City Elementary School System," *Yale Economic Essays,* Vol. 8, No. 1 (Spring 1968).

Kiesling, Herbert J., "Measuring a Local Government Service: A Study of School Districts in New York State," *Review of Economics and Statistics,* Vol. XLIX, No. 3 (August 1967).

2

Levin, Henry M., *Recruiting Teachers for Large City Schools* (New York: Charles Merrill and Sons, 1970).

Mayeske, George W., *et al.*, Technical Note Number 61, *Correlational and Regression Analyses of Differences Between the Achievement Levels of Ninth Grade Schools from the Educational Opportunities Survey* (Washington, D.C.: U.S. Office of Education, National Center for Educational Statistics, March 11, 1968).

Mollenkopf, William G., and Melville, S. Donald, "A Study of Secondary School Characteristics as Related to Test Scores," Research Bulletin 56-6 (Princeton: Educational Testing Service, 1956), mimeographed.

Mort, Paul R., "Cost Quality Relationships in Education," *Problems and Issues in Public School Finance*, edited by Johns, R. L., and Morphet, Edgar L. (New York: National Conference of Professors of Educational Administration, 1952).

Raymond, Richard, "Determinants of the Quality of Primary and Secondary Public Education in West Virginia," *Journal of Human Resources*, Vol. 3, No. 4 (Fall 1968).

Ribich, Thomas I., *Education and Poverty* (Washington, D.C.: Brookings Institution, 1968).

Shaycoft, Marion F., *The High School Years: Growth in Cognitive Skills* (Pittsburgh: American Institute for Research and School of Education, University of Pittsburgh, 1967).

Thomas, J. Alan, "Efficiency in Education: A Study of the Relationship between Selected Inputs and Mean Test Scores in a Sample of Senior High Schools," (unpublished Ph.D. dissertation, Stanford University: School of Education, 1962).

Warner, W. Lloyd, Meeker, Marcia, and Eels, Kenneth, *Social Class in America* (Chicago: Science Research Associates, Inc., 1949).

Weikart, David P., *Preschool Intervention: A preliminary Report of the Perry Preschool Project* (Ann Arbor, Michigan: Campus Publishers, 1967).

Welch, Finis, "Measurement of the Quality of Schooling," *American Economic Review*, Vol. LVI, No. 2 (May 1966).

Werts, Charles E., and Linn, Robert L., "Analyzing School Effects: How to Use the Same Data to Support Different Hypotheses," *American Educational Research Journal*, Vol. VI, No. 3 (May 1969).

Wilson, Alan B., "Residential Segregation of Social Classes and Aspirations of High School Boys," *American Sociological Review*, Vol. 24 (1959).

Chapter 3

A NEW MODEL OF SCHOOL EFFECTIVENESS

Henry M. Levin

The subject of how schools affect the development of youngsters has been under intensive study for at least 50 years. In most cases the unit of analysis has been the classroom where attempts are made to relate differences in environmental and interaction variables to differences in student performance. The usual approach has been to set up experimental and control groups, to apply the "treatment" to the experimental one, and to look for significant differences in outcomes between the two groups. Unfortunately, the extensive research utilizing this methodology has not come up with a reasonably consistent and reproducible set of findings on how differences in schools create differences in human development.

Certainly one of the reasons for the inability of these experiments to provide useful conclusions is the assumption of *ceteris paribus,* i.e., all other things being equal between control and experimental groups. Rather, the complexity of the world within which education takes place suggests that observed similarities between control and experimental groups on one or two dimensions is not adequate for the *ceteris paribus* assumption. Many influences must be accounted for in seeking the determinants of scholastic achievement, attitude formation, and so on.

In the last decade a number of studies have attempted to go beyond the standard type of educational experiment by using large-scale multivariate statistical models to account for many more variables than could be included in the typical control group/experimental group comparison. These studies have related the achievement of students to variables reflecting the student's race, socioeconomic status, teacher, and other school variables, as well as the characteristics of fellow students. The rather consistent set of findings emerging from these studies suggests that three measured factors are significantly related to student academic achievement: (1) race of student, (2) socioeconomic status of the student, and (3) characteristics of his teachers.[1]

Generally these endeavors have utilized survey data on student achievement, socioeconomic backgrounds, and school resources to explain variance in student achievement. Typically, their findings are based upon fitting a linear regression via the ordinary least-squares criterion for the following formulation.

$$A_{it} = f(B_{it}, S_{it}, O_{it})$$

where A_{it} is the standardized achievement score of the ith student at time t; B_{it} represents a vector of family background characteristics at time t; S_{it} represents a vector of school resources such as teacher characteristics, facilities, student environment created by peers, and so on at time t; and O_{it} represents community and other characteristics that might affect achievement. These attempts might be conceived of broadly as attempts to estimate educational production functions. That is, studies of the educational production process are analogous to the econometric effort of estimating production processes in other industries.[2] While it is not the purpose of this study to review all of the properties of educational production functions and the problems encountered in estimating them, it is useful to discuss briefly a few of these.

The Focus on a Single Output

Most studies of the educational production function have used standardized achievement scores as the output of the process. Yet, schools are expected to produce many outcomes in addition to increasing academic achievement.[3] The formation of a variety of attitudes and skills as well as many social externalities are attributed to the schools.[4] An empirical analysis of educational production that considers only one output ignores these other outcomes. Only if these other outcomes are produced in fixed proportion to the output under scrutiny does no problem arise in focusing on a single output such as standardized achievement.[5]

Ideally, the estimation of the educational production process should be based upon total educational output. That is, in some way we would want to weight the outputs produced by some common factor (utilities, votes, social values) in order to obtain a total index of output. Multiproduct firms that sell their outputs in the marketplace are able to obtain such a measure by using prices as weights to obtain a monetary value for total product. Unfortunately, we can neither measure all of the outputs that schools are supposed to produce nor do we possess a yardstick or "numeraire" to put them into an index of output.

This focus on achievement scores as the single measure of school output creates at least two problems in measuring the educational production process.

First, the single focus on achievement limits the usefulness of educational production studies to providing insights for only one dimension of school output. The efficient ordering of inputs for producing achievement may be exceedingly inefficient for increasing student motivation, efficacy, imagination, and other desirable outcomes. This study will attempt to partly reconcile this problem by considering relationships among educational inputs and several outputs.

Second, estimates of the educational production process will underestimate the relation between any single output and school resources as long as priorities for that output vary among schools. To take an extreme case, academic high schools tend to emphasize language skills much more heavily than do vocational high schools. Accordingly, equal resources devoted to both groups of schools, *ceteris paribus,* would likely have a greater impact on verbal achievement among the academic students than the vocational ones.

This relationship is further confounded if the priorities of schools vary according to the socioeconomic composition of their student bodies. Certainly, the middle class schools are generally more academically oriented in a college-preparatory sense than are the lower class schools which seem to emphasize more heavily the general or job-oriented curriculums. In such a case the socioeconomic background variables of the students act as a proxy for the emphasis on academic skills relative to other school goals, and their statistical importance in "producing" academic achievement scores will be overstated while the impact of school resources will be understated.

Educational Production Theory and the Meaning of Production Data

Estimates of production functions in other industries are based upon the assumption that firms are maximizing output for any set of inputs; that is, firms are assumed to be technically efficient. Only under these conditions will estimates relating inputs to output reflect the most efficient way of producing that output.

In order to satisfy that assumption there are at least three general conditions that must be presumed: (1) the firm has knowledge of the relevant production set; (2) the firm has discretion over the way in which inputs are used; and (3) there is an effective incentive that spurs the firm to apply its knowledge of the production set and its ability to combine inputs into maximizing output for any set of physical inputs. Under these conditions the observed production data depict the production frontier, the largest output attainable for each set of inputs. Whether these are valid presumptions for private firms may be open to question, but they are clearly inappropriate ones for the schools.[6] There is no basis for asserting that educational decision-makers know their relevant production sets or that they have a great deal of discretion over how their inputs are used. The present organization of school inputs tends to be based on sacrosanct traditions rather than management discretion. Finally, the incentives of the marketplace that spur firms to be technically (and allocatively) efficient—profits, sales, and so on—are conspicuously absent from the educational scene. In particular, there is no evidence that educational firms such as schools and school districts maximize standardized achievement. Thus, at best the observations on inputs and outcomes represent average ones under the present state of operations, not maximum or technically efficient ones.

Moreover, the lack of knowledge on the relevant production set means that one cannot specify with reasonable accuracy the inputs germane to any particular output. Specification of the educational production model must depend more on intuition and hunch than on a body of well-developed behavioral theory. That is, there is no well-validated theory of learning on human development which can be used as a guide in specifying inputs and the general functional relationships between inputs and outputs. In the absence of such a foundation, much of the early work in estimating educational production relations has necessarily involved a hunting expedition into the deep entangled forest of possible educational influences. The problem with such an expedition is that we have been like hunters shooting at anything that moved since we have had no clear picture of the animals we wanted to collect.

A second and related problem is that even when we do know what kind of conceptual animal we wish to bag, we do not know how or where to capture it. Clearly, innate intelligence should be considered as an input when attempting to estimate the educational production function for achievement. Yet, like the mythical unicorn, much has been written about innate intelligence, no one has ever seen one. That is, we have no way of measuring this important determinant of educational outcomes. Moreover, measures of teacher proficiency or other school inputs are not available. Rather we must use such conventional indicators as teacher experience, degree level, number of books in the library, and so on in the hope that we are capturing some of the actual influences of which we are unaware or which we are unable to measure adequately.

The result of both not knowing how education is produced and not being able to measure many of the inputs suggests a high probability of bias in the estimates of the production coefficients. The exclusion of variables that belong in the equation as well as the inclusion of erroneous variables all lead to such biases.[7] Moreover, the fitting of such data to a linear function can also result in specification biases in a world that is characterized by nonlinearities. All of the empirical studies of the educational production process are prime candidates for such biases.

Data Refinement

Perhaps it is useful to divide data problems into two types: intransigent and remediable. In actuality this dichotomy is a state-of-the-art distinction rather than one which is in the stars. At a future time, intransigent difficulties may be alleviated by greater knowledge of the phenomenon or by better measurement techniques. Examples of the former problem are our inability to measure innate abilities. As we noted above, the omission of such a variable is likely to induce a bias in our estimates. In such a case it is important that we explore the biases from not including such a measure in the specification of our production model, but we can do little beyond this.

On the other hand, data deficiencies arise that are partly or fully remediable. For example, a needed item is sufficiently measurable, but it was omitted from the survey on which the production estimates will be made. In such a case, one can attempt to find a close proxy among the existing information source or one can resurvey to obtain the missing item. The latter alternative is time consuming and costly, so it is often the former course of action that is taken. Yet, the use of a proxy or surrogate piece of information is subject to the vagaries of interpretation, and its use may create more problems than it solves.[8] In many cases it may be wise to acknowledge the omission and to speculate on the resulting bias rather than to use a questionable proxy.

Yet, in all too many instances data problems are remediable, and in those cases the information should be refined to more closely approximate the concept which they are expected to represent. Most studies examining the educational production process have used school data for each student whether the student had actually attended the school in the past or whether he hadn't. For example, the Equality of Educational Opportunity

(EEO) survey was undertaken in September-October of the 1965-66 school year. Clearly the relevant school data for each child are those pertaining to the schools that he actually attended, and in many cases the school that he was attending in 1965-66 was different from those that he had previously attended. That is, the high rate of residential mobility is translated into school mobility, and present school factors may be erroneous measures for actual school characteristics unless some data refinement is attempted.[9]

To the degree that the school factors used in the analysis are spurious ones, the estimated effect of them on achievement will be biased downward.[10] Unfortunately, this problem pervades the EEO work as well as its reanalysis, and the problem is more serious among the analyses for blacks and other minorities than for whites because of the higher mobility factor among the former groups. One way of correcting for this source of error is to include in the sample only those students who had received all of their education in the schools which they were currently attending. That is the approach taken in this study. Another possibility is that of obtaining historical data on all of the schools that the students attended. Given the fact that much school mobility is among school districts and States, this task may be beyond the realm of practicality.

Other data problems that are remediable are those resulting from missing observations of items for particular students. The EEO survey suffered particularly from these hindrances.[11] There are many ways of handling this problem, but ignoring it is clearly not one of them.[12] A final difficulty that characterizes the data sets used for measuring educational production is the interdependence among the so-called explanatory variables. In general, a child's home background and his school are highly correlated in that higher socioeconomic status children attend schools with greater resource endowments. This factor has prevented many studies from obtaining reliable estimates of the separate effects of school and background characteristics on achievement.[13] One way of circumventing this difficulty is to carry out the analysis for stratified subsamples of students with homogeneous socioeconomic backgrounds.[14]

Purpose of This Study

While we have noted some of the problems that arise in applying econometric analysis of production to the schools, this study will not make the heroic claim of having avoided such pitfalls. Rather, this effort addresses itself to moving toward estimating a model of the schools that more nearly mirrors what we know of the educational process. Indeed, we will proceed in the following way: First, we will posit a model of the schools and compare it with the more traditional formulation; second, we will discuss the data that will be used to estimate the structure of the model; third, we will review the estimation procedure and results; and finally, we will discuss the implications.

Specification of the Model

Most studies of educational production have not attempted to specify in a systematic way the particular formulation of how schools affect achievement. Rather, they have taken a set of school and student background factors and related them statistically to achievement without discussing the underlying behavioral assumptions implied by their work. One exception has been an important study by Eric Hanushek that did posit a more concrete model of achievement.[15] The following formulation is based upon Hanushek's foundation.

Assume that we wish to examine the determinants of student achievement at a point in time. Clearly, that achievement level is related not only to the present influences that operate on that student, but also to past ones. That is, from the time a child is conceived various environmental characteristics combine with his innate characteristics to mold his behavior. More specifically, a child's achievement performance is determined by the cumulative amounts of "capital" embodied in him by his family, his school, his community, and peers as well as his innate traits. The greater the amount and the quality of investment from each of these sources, the higher will be the student's achievement level. Thus, a student's academic performance is viewed to be a function of the amount of different kinds of capital embodied in him.

The general formulation of the capital embodiment model is as follows:

$$A_{it} = g\left[F_{i(t)}, S_{i(t)}, P_{i(t)}, O_{i(t)}, I_{it}\right] \tag{1}$$

where the i subscript refers to the ith student? the t subscript refers to time period t; and the t subscript in parentheses (t) refers to being cumulative to time period t. Thus:

A_{it} a vector of educational outcomes for the ithe student at time t.
F_{it} a vector of individual and family background characteristics cumulative to time t.
$S_{i(t)}$ = a vector of school inputs relevant to the ith student cumulative to time t.
$P_{i(t)}$ = a vector of peer or fellow student characteristics cumulative to t.
$O_{i(t)}$ = a vector of other external influences (community, etc. . .) relevant to the ith student cumulative to t.
I_{it} = a vector of initial or innate endowment of the ith student at t.

It is assumed that g' is positive for all these arguments or that the marginal product of additional capital embodiment from any one of the five sources has a positive effect on student educational outcome.[16]

This formulation reflects the well-accepted concept that a child receives his educational investment from several sources in addition to the school. For example, the family provides a material, intellectual, and emotional environment which contributes to the child's performance level. Likewise, the school, peer groups, and community affect both learning and emotional behavior of students. Yet, in order to estimate these effects, one must take this general formulation and make it more specific.

Suppose we wish to follow the examples of other researchers by estimating a production function for achievement. Again, we can view a student's level of achievement on a verbal test, for example, as a function of his capital embodiment from several sources as well as his innate traits. But, in addition to these sources of capital embodiment, his educational achievement at a point in time is likely to be related to his educational attitudes and his parents' educational attitudes. More specifically, we might postulate that:

$$(2) \qquad A_{1_{it}} = g[F_{i(t)}, S_{i(t)}, P_{i(t)}, O_{i(t)}, I_{i(t)}, A_{2it}, A_{3it}, A_4]$$

where

$A_{1_{it}}$ = the achievement level of the ith student at t.

$F_{i(t)}, S_{i(t)}, P_{i(t)}$, and $O_{i(t)}$ are as previously defined.

$A_{2_{it}}$ = a measure of the student's sense of efficacy or fate control at t.

$A_{3_{it}}$ = a measure of educational motivation of the ith student at t.

$A_{4_{it}}$ = parents' educational expectations for the ith student at t.

That is, we would expect student achievement to be higher the greater his sense of efficacy, his educational motivation, and his parents' expectations, *ceteris paribus.* By efficacy we refer to the student's feeling that he has a measure of control over his destiny, that it does not depend strictly on chance. Educational motivation refers to the desire to succeed in an educational sense (for example, the desire to get good grades and to attain additional schooling). Parents' educational expectations might be viewed as how well the parents expect the child to perform by educational criteria.

But these three variables are of more than passing interest because *not only do they affect achievement levels, but they themselves are affected by achievement.* This raises the question of whether a single equation is adequate for estimating educational production, even when one is concerned with only a single measure of output such as achievement. That is, the single equation model tacitly assumes that each of the explanatory variables is determined outside of the system; that is they are exogenous. In other words, the explanatory variables influence the level of student achievement, but student achievement is assumed not to influence the so-called explanatory variables.

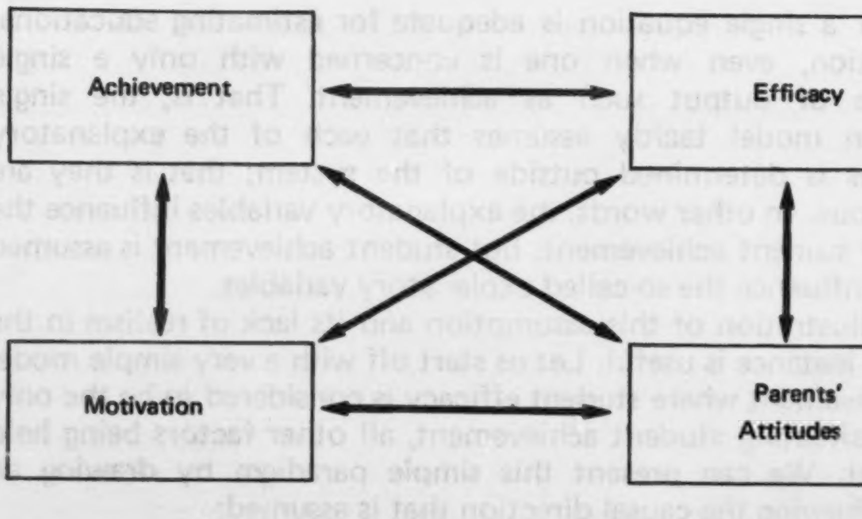
An illustration of this assumption and its lack of realism in the present instance is useful. Let us start off with a very simple model of achievement where student efficacy is considered to be the only factor affecting student achievement, all other factors being held constant. We can present this simple paradigm by drawing an arrow showing the causal direction that is assumed:

Student Achievement ← Student Efficacy

This simple depiction suggests that student achievement is greater when the level of efficacy is higher. In process terms, students who believe that they have a measure of control over their achievement level are more likely to try to do well than students who believe that it all depends upon luck. But it is probably also true that the higher the level of his achievement, the higher the level of his efficacy. That is, by doing well, his sense of fate control is enhanced or reinforced because his efforts can really make a difference in his achievement. Thus, achievement stimulates efficacy and efficacy stimulates achievement as depicted below:

Student Achievement ↔ Student Efficacy

Moreover, the other attitudinal variables that influence such school outputs as standardized achievement performance are also influenced themselves by student achievement and by each other. For example, parents' educational expectations for a student will affect the student's performance level; but the student's performance level will also affect the parent's educational aspirations for him. Most parents will expect less from a child who has consistently low test scores and grades than one who has higher levels of both attributes. The same is probably true of teacher expectations for pupil progress. In summary, many crucial variables in the educational process interact in such a way that we cannot take their levels as given in order to predict other factors. Rather both explanatory variables and those which we wish to explain are interdependent, and their values must be solved simultaneously in order to obtain unbiased estimates of their effects. That is, the following relationship exists in concept. In this particular system, everything de-



pends upon everything else, so that complete simultaneity exists. Every one of the variables is linked by a double arrow to every other variable. In actuality the simultaneity may be complete or partial, but in either case the ordinary least-squares solution of equation (2) will lead to biased and inconsistent estimates.[17] Rather, we must estimate the full set of equations representing the simultaneous equations system.

The following formulation describes the simultaneous equation model:

$$(3) \quad A_{1_{it}} = g_1\left[F_{1_{i(t)}}, S_{1_{i(t)}}, P_{1_{i(t)}}, O_{1_{i(t)}}, I_{1_{it}}, A_{2_{it}}, A_{3_{it}}, A_{4_{it}}\right]$$

$$(4) \quad A_{2_{it}} = g_2\left[F_{2_{i(t)}}, S_{2_{i(t)}}, P_{2_{i(t)}}, O_{2_{i(t)}}, I_{2_{it}}, A_{1_{it}}, A_{3_{it}}, A_{4_{it}}\right]$$

$$(5) \quad A_{3_{it}} = g_3\left[F_{3_{i(t)}}, S_{3_{i(t)}}, P_{3_{i(t)}}, O_{3_{i(t)}}, I_{3_{it}}, A_{1_{it}}, A_{2_{it}}, A_{4_{it}}\right]$$

$$(6) \quad A_{4_{it}} = g_4\left[F_{4_{i(t)}}, S_{4_{i(t)}}, P_{4_{i(t)}}, O_{4_{i(t)}}, I_{4_{it}}, A_{1_{it}}, A_{2_{it}}, A_{3_{it}}\right]$$

In this system there exists an equation for each of the endogenous variables. Two characteristics of the sytem are of immediate importance. First, the solution of the system depends upon its identifiability. In general, proper identification requires that there be as many equations as endogenous variables and that all variables are not present in all relations.[18]

In this regard, it should be noted that the specification of each of the exogenous variables is unique in each relation. That is, it is reasonable to believe that different family factors, school factors, innate characteristics, and so on, affect achievement $\left(A_{1_{it}}\right)$ than affect the other endogenous variables $\left(A_{2_{it}}\right)$, $\left(A_{3_{it}}\right)$, and $\left(A_{4_{it}}\right)$. Accordingly, $F_{1_{i(t)}}$ is considered to be a different vector of family influences than $F_{2_{i(t)}}$, $F_{3_{i(t)}}$, and $F_{4_{i(t)}}$.[19]

The potential uniqueness of $S_{i(t)}$ for each equation is also represented by the appropriate subscript as well as the uniqueness of the other vectors. It is particularly useful if we can distinguish between school characteristics that relate to achievement $\left(A_{1_{it}}\right)$ and those that relate to student and parental attitudes.

A second characteristic of the system represented by equations (3), (4), (5), and (6) is that each of the endogenous variables represents an *output* of the educational process as well as an input into it. Just as schools are expected to increase achievement, they are also expected to contribute to such attitudes as efficacy and motivation. Thus, we can evaluate the system for each of several outcomes rather than restricting ourselves only to the analysis of student achievement.[20] The system of equations allows us to solve for student efficacy, student motivation, and parents' educational expectations as well as student achievement.

Estimating the Equations

The data used to estimate this system were derived from the Equal Opportunity Survey on which the Coleman Report was based. The sample is composed of sixth-grade students in a large eastern city who had attended only the school in which they were enrolled at the time of the survey, 1965-66.[21] Teacher characteristics are based upon averages for all of the teachers in each school who were teaching in grades 3-5. These averages were intended to reflect the teacher characteristics that had influenced student behavior up to the time of the survey. Since family background characteristics and other educational influences were measured only at a point in time, it is tacitly assumed that these measures bear a constant relation to the stock of capital embodied in each child from these sources. That is, it is assumed that the values of those inputs cumulative to time t bear a constant relation to the flow of inputs observed at time t.

While all of the equations specify innate traits as exogenous variables, we do not possess measures of I_{it}. That is, our statistical model does not include the I_{it} vectors despite the fact that they belong in the system, *a priori.* It is important to speculate on the expected bias in the estimates of the other parameters, if the students' innate traits are not included in the equations. In general, those variables that are correlated with the omitted one will be biased upwards.[22]

It is probably reasonable to assume that innate traits have at least some component that is reflected in the vector of family background characteristics.[23] Even if one minimizes the possible genetic relation between parental traits and the child's innate characteristics, there are other possible linkages. In particular, the child drawn from lower origins is a more likely candidate for prenatal protein starvation, a factor which may limit his innate potential.[24] The result of the probable association between family background characteristics and student's innate traits is that the effect of the $F_{i(t)}$ vector on achievement (and perhaps on other outcomes) will be overstated. That is, family background characteristics will be biased upwards to the extent of their covariance with the missing variable, innate characteristics. In general, it is reasonable to conclude that all of the studies that have tried to explain the determinants of scholastic achievement have overstated the effects of family background by omitting measures of innate traits.

Some Results

What follows are some estimates of a simultaneous equation system similar to that posited above. The particular sample in this analysis consists of almost 600 white students attending some 36 schools in Eastmet City. The basis on which particular variables were chosen to enter the relation was based partially on *a priori* judgment, partially on statistical tests of significance, and partially on the quality of the measures.

On the basis of over 100 items of information that we distilled from the original survey data, we chose those variables that might be expected, logically, to enter into each relation. As an example, the quality of library services as represented by library books per student might reasonably be expected to affect the student's achievement level; yet, one would be hard pressed to discern a direct relationship between student's and parents' attitudes and library books. Accordingly, the library measure was specified only in the achievement equation. Likewise, such information as teacher's salary is reflected in the teacher characteristics that the salaries purchase.[25]

Some items that were entered showed statistical relationships that were so nearly random that they were eliminated from subsequent equations. Whether the lack of a statistical association was due to their poor measurement or their misspecification cannot be determined *a priori*. What follows is a set of estimates that must be judged only for their heuristic values. That is, alternative specifications are equally plausible, and the grounds for specification biases are substantial.[26] Further refinement of the data and the specifications are undoubtedly necessary before firm policy influences can be drawn.

Table 1 shows the list of all variables included in the estimates; and tables 2, 3, 4, and 5 show estimates of the equations for verbal

TABLE 1

List of Variables in Simultaneous Equations System

Name of Variable	Measure of	Coding
Verbal Score	Student Performance	Raw Score
Student's Attitude	Efficacy	Index compiled from questions 33-40 in the Sixth Grade Student Questionnaire of the Equal Opportunity Survey. (e.g., I can do many things well. Well No Not Sure I sometimes feel I just can't learn. Yes No The higher the value of the index, the greater the perceived efficacy of the student.)
Parents' Attitude	Educational Expectations of Parents	Index based upon three questions: (1) How good a student does your mother want you to be? (2) How good a student does your father want you to be? (3) Did anyone at home read to you when you were small, before you started school? (and how often?)
Grade Aspiration	Student Motivation	Grade level the student wishes to complete
Sex	Male-Female Differences	Male = 0 Female = 1
Age	Overage for Grade	Age 12 or over = 1 Less than 12 = 0
Possessions in Student's Home	Family Background (Socioeconomic Status)	Index of possessions: Yes = 1 No = 0 for each; Index is sum. television telephone dictionary encyclopedia automobile daily newspaper record player refrigerator vacuum cleaner
Family Size	Family Background	Number of people in home
Identity of Person Serving as Mother	Family Background	Real mother at home = 0 Real mother not living at home = 1 Surrogate mother = 2
Identity of Person Serving as Father	Family Background	Real father at home = 0 Real father not living at home = 1 Surrogate father = 2
Father's Education	Family Background	Number of years of school attained

TABLE 1—*(Continued)*

Name of Variable	Measure of	Coding
Mother's Employment Status	Family Background	Has job = 1 No job = 0
Attended Kindergarten	Family Background	Yes = 1 No = 0
Teacher's Verbal Score	Teacher Quality	Raw score on vocabulary test
Teacher's Parents' Income	Teacher Socioeconomic Status	Father's occupation scaled according to income (1000's of dollars)
Teacher Experience	Teacher Quality	Number of years of full-time experience
Teacher's Undergraduate Institution	Teacher Quality	University or college = 3 Teacher institution = 1
Satisfaction with Present School	Teacher's Attitude	Satisfied = 3 Maybe prefers another school = 2 Prefers another school = 1
Percent of White Students	Student Body	Percentage estimated by teachers
Teacher Turnover	School	Proportion of teachers who left in previous year for reasons other than death or illness
Library Volumes Per Student	School Facilities	Number of volumes divided by school enrollment

NOTE: All data are taken from the Equal Opportunity Survey for Eastmet City. The survey instruments are found in James S. Coleman *et al., Equality of Educational Opportunity* (Washington, D.C.: U.S. Government Printing Office, 1966).

TABLE 2

Estimates of Verbal Score Equations for White Sixth Graders in Eastmet City (t values in parentheses)

	Ordinary Least Squares	Two Stage Least Squares	Reduced Form
Student's Attitude	0.641 (4.88)	2.649 (1.72)	
Grade Aspiration	0.921 (5.21)	0.591 (0.53)	
Parents' Attitude	0.605 (2.81)	0.873 (0.74)	
Sex	0.616 (1.06)	–0.571 (0.49)	0.817
Age	–6.099 (4.26)	–5.513 (2.78)	–6.010
Possessions	0.990 (3.84)	0.521 (1.05)	1.229
Family Size	–0.330 (2.14)	–0.036 (0.12)	–0.552
Identity of Mother	-----	-----	–0.433
Identity of Father	-----	-----	–0.327
Father's Education	0.243 (2.10)	0.026 (0.12)	0.273
Mother's Employment	-----	-----	–0.509
Attended Kindergarten	1.520 (1.73)	1.768 (1.32)	2.372
Teacher's Verbal Score	0.332 (1.61)	0.220 (0.84)	0.250
Teacher's Parent's Income	-----	-----	–0.118
Teacher Experience	0.751 (8.77)	0.694 (5.28)	0.787
Teacher Undergraduate Institution	6.547 (2.66)	5.833 (1.94)	6.525
Satisfaction with Present School	1.201 (0.90)	1.658 (0.86)	1.960
Percent of White Students	-----	-----	–0.047
Teacher Turnover	–0.054 (0.61)	0.044 (0.34)	–0.101
Library Volumes Per Student	0.562 (1.82)	0.498 (1.31)	0.565
Constant Term	–23.94	–29.75	–7.902
$\bar{R}^2$	.53	.34	

TABLE 3

Estimates of Student Attitude Equations for White Sixth Graders in Eastmet City
(t values in parentheses)

	Ordinary Least Squares	Two Stage Least Squares	Reduced Form
Verbal Score	0.061 (5.54)	0.052 (2.03)	
Parents' Attitude	0.112 (1.69)	0.042 (0.15)	
Sex	0.560 (3.15)	0.557 (3.08)	0.577
Age	0.241 (0.54)	0.135 (0.27)	−0.015
Possessions	0.107 (1.39)	0.143 (1.29)	0.174
Family Size	−0.108 (2.30)	−0.124 (2.05)	−0.138
Identity of Mother	-----	-----	−0.011
Identity of Father	−0.082 (1.30)	−.092 (1.36)	−0.100
Father's Education	0.070 (2.02)	0.081 (1.88)	0.088
Mother's Employment	−0.318 (1.58)	−0.307 (1.44)	−0.320
Attended Kindergarten	-----	-----	0.059
Teacher's Verbal Score	-----	-----	0.006
Teacher's Parents' Income	-----	-----	−0.003
Teacher Experience	-----	-----	0.163
Teacher Undergraduate Institution	-----	-----	0.020
Satisfaction with Present School	−0.163 (0.42)	−0.129 (0.33)	−0.089
Percent of White Students	-----	-----	−0.001
Teacher Turnover	−0.047 (2.70)	−0.048 (2.73)	−0.051
Library Volumes Per Student	5.132	5.330	5.132
$\bar{R}^2$	.19	.19	

TABLE 4

Estimates of Grade Aspiration Equations for White Sixth Graders in Eastmet City (t values in parentheses)

	Ordinary Least Squares	Two Stage Least Squares	Reduced Form
Verbal Score	.0557 (6.75)	.0876 (4.18)	
Parents' Attitude	.0372 (0.75)	–0.391 (1.46)	
Sex	–0.111 (1.84)	–0.192 (1.30)	–0.077
Age	–0.351 (1.05)	–0.243 (0.63)	–0.772
Possessions	0.052 (0.87)	.074 (0.85)	0.092
Family Size	–0.057 (1.64)	–0.077 (1.62)	–0.079
Identity of Mother	–0.223 (2.35)	–0.310 (2.62)	–0.227
Identity of Father	–0.056 (1.11)	–0.560 (1.03)	–0.077
Father's Education	-----	-----	0.024
Mother's Employment	0.282 (1.89)	0.401 (2.34)	0.279
Attended Kindergarten	0.644 (3.20)	0.547 (2.47)	0.756
Teacher's Verbal Score	-----	-----	0.022
Teacher's Parents' Income	–0.0005 (0.38)	–0.176 (1.15)	–0.186
Teacher Experience	-----	-----	0.069
Teacher Undergraduate Institution	–0.460 (1.08)	–0.135 (0.28)	0.439
Satisfaction with Present School	0.785 (2.56)	0.693 (2.80)	0.866
Percent of White Students	-----	-----	0.021
Teacher Turnover	-----	-----	–0.005
Library Volumes Per Student	-----	-----	0.050
Constant Term	9.174	10.900	8.850
$\bar{R}^2$	.26	.15	

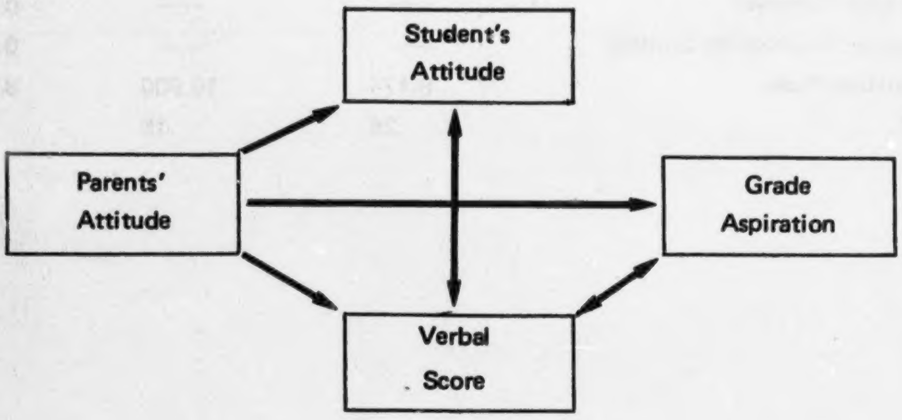
TABLE 5

Estimate of Parent's Attitude Equation for White Sixth Graders in Eastmet City (t values in parentheses)

	Ordinary Least Squares
Sex	−0.110 (1.00)
Possessions	0.218 (4.84)
Family Size	−0.119 (4.14)
Identity of Mother	−0.309 (4.36)
Identity of Father	−0.018 (0.44)
Mother's Employment	0.198 (1.59)
Percent of White Students	−0.065 (2.11)
Teacher's Turnover	−0.009 (.89)
Constant Term	3.465
$\overline{R}^2$	.13

score, student's attitude, grade aspiration, and parents' attitude, respectively. The sample comprises 597 white students in the sixth grade of Eastmet City in the fall of 1965.

Before interpreting the results, it is important to note that the statistical model used here differs slightly from that shown in equations (3), (4), (5), and (6) in that only the first three equations are estimated simultaneously. That is, the fourth equation is estimated by ordinary least-squares, and it bears a recursive relation to the rest of the model. The figure that follows illustrates this property as well as the simultaneous relationships estimated among the other equations. The system is overidentified *a priori*



because the endogenous variables are not common to each of the three simultaneous equations.

Two-stage least-squares was used for the three simultaneous equations. Each of the tables for the equations on verbal score, student's attitude, and grade aspiration show an ordinary least-squares estimate, a two-stage least-squares (simultaneous equations) estimate, and a reduced form. The latter is obtained by solving the simultaneous equations system via algebraic substitution.[27]

Some Interpretations

The interpretations that are given here are highly speculative. They are offered only as illustrations of the properties of the model. Further testing of the structure and improved data are necessary to confirm results reported here. Accordingly, the interpretation of the findings is not an attempt to be exhaustive as much as it is an effort to show how this approach might be used ultimately to examine various hypotheses.

Verbal Score

The variables entering the verbal score equation were selected as being representative of the different vectors in equation (3) with the obvious omission of innate traits. Such conventional teacher's characteristics as degree level showed no significant relation with student verbal score, although teacher's experience appears to be strongly related in this sample.

It is especially instructive to compare the ordinary least-squares estimates (which do not take account of the simultaneity) with the two-stage estimates (which do take account of it). In this way we can note some of the biases in interpretation that might arise from the usual ordinary least-squares estimates. In particular it appears that the direct effect of several family background characteristics on verbal achievement is overstated substantially in the single equation (OLS) estimate. For example, the coefficient for family size is only one-tenth as large in the TSLS estimate as the OLS one. This suggests that the large observed negative relation between family size and achievement in the ordinary least-squares formulation should not be interpreted as a direct effect, but one that works through an intervening variable, student's attitude. The much larger coefficient for student's attitude in the TSLS estimate in combination with the great decline in the family size coefficient in the simultaneous-equations formulation indicates that students from larger families probably have lower verbal scores because of their poorer attitudes rather than because of an inextricable link between family size and other background characteristics on the one hand and achievement on the other. The existence of this phenomenon is also supported by the smaller coefficients in the TSLS estimate for such socioeconomic factors as father's education and possessions.

The possible significance of these findings is that educational programs that focus on student attitudes may be able to compensate for "disadvantages" in socioeconomic background. Indeed, this tentative interpretation argues against the simplistic observations of some social philosophers that educational programs cannot compensate for such background deficiences as low socioeconomic status—since these background factors now appear to have much of their direct effects not on achievement, but on attitude and *through* attitude, on achievement. Successful efforts to change student attitudes, therefore, might be used to offset "deleterious" background conditions.

In this vein it is also interesting to note the reversal of sign for the sex variable between the OLS and TSLS estimates. In the OLS formulation females show higher verbal scores than males, while in the TSLS they show lower scores. Again, it appears that the higher verbal scores of females are more likely attributable to a higher sense of efficacy rather than to any direct sex-achievement effect. This is confirmed by the strong, positive coefficient for females in the student attitude equation in table 3. It is also supported by the well established view that schools represent feminizing influences, receptive to girls and hostile to boys. Under such conditions one would expect females to have greater efficacy and through efficacy, greater achievement.[28]

The reduced form equation shows all of the system's influences on verbal score—whether directly through the verbal score equation or indirectly through students' attitudes, grade aspiration, or parents' attitudes. On balance, sex is positively related to verbal score. Those variables that affect attitudes and grade aspiration directly are shown to affect verbal score because attitudes and grade aspiration affect verbal score. Thus, while the identity of the mother showed no significant direct relation with verbal achievement it does show a negative influence of a maternal substitute in the reduced form because of its direct negative relation on student grade aspiration. The same is true of father's identity which shows a direct negative effect of a father surrogate on student's attitude and thus on indirect effect in the reduced form on verbal score.

Other Equations

Table 3 presents comparable equations for student's attitude and table 4 shows them for grade aspiration. Because of the tentative nature of the findings at this stage of the art, we will not detail all of these results. Rather, we will focus on a pattern that is of general interest. In particular, it appears that when the mother has a job, the child's grade aspiration is higher (table 4), but his efficacy or attitude is lower (table 3). Even in the reduced forms of these two equations, the differences in sign prevail, and in the reduced form on verbal score (table 2) a child whose mother works shows a lower test performance ostensibly because of the effect of his mother's employment on his own efficacy.

The findings in these tables are pregnant with suggestions, and it is interesting to speculate on their meaning. Yet we must caution against any final interpretation until improved measurement and replication of the model confirm the observed patterns. Accordingly, it is best to summarize where this excursion has taken us.

A Summary

In this paper an analogy between the economist's concept of an educational production function has been outlined. The problems of estimating the same have been emphasized. Despite these obstacles, the importance of knowing the production relationships in the educational sector has stimulated much recent research. The effort presented in this paper is an extension of this research by positing a simultaneous-equations approach for viewing the educational process. It appears that the properties of a simultaneous-equations system mirror the world more closely than the single-equation approaches that are presently being used. Further developments in this direction are proceeding, and it is hoped that before long, we can obtain a reasonably reliable set of estimates of school effectiveness by using this technique.

Acknowledgments

The analysis in this paper was drawn from a larger study which is being authored jointly with Stephan Michelson. Research support has been provided by the Stanford Center for Research and Development in Teaching, and the editing of the data was done at the Brookings Institution. The Center for Educational Policy Research at Harvard supported some of the computational costs. In addition to Stephan Michelson, the author is indebted to Randall Weiss for his research assistance and for his thoughtful contributions. Emily Andrews assisted in the final preparation of this paper.

Footnotes

[1] See the survey of these studies in James W. Guthrie, George B. Kleindorfer, Henry M. Levin, and Robert Stout, *Schools and Inequality: A Study of the Relationships between Social Status, School Services, and Post-School Opportunity in the State of Michigan,* a report prepared for the National Urban Coalition, Washington, D.C. (mimeo, September 1969).

[2] For a survey of econometric work on production functions see A. A. Walters, "Production and Cost Functions: An Econometric Survey," *Econometrica,* Vol. 31, Nos. 1-2 (January-April, 1963), pp. 1-66. The most comprehensive work on educational applications is Samuel S. Bowles, "Towards an Educational Production Function." A paper prepared for the Conference on Research in Income and Wealth (Madison, Wis., November 1968), mimeo. The theory of production can be found in any basic text on microeconomics. See for example, William J. Baumol, *Economic Theory and Operations Analysis* (Englewood Cliffs, N.J.: Prentice - Hall, 1963), Chapter 11.

[3] For classifications of these, see Benjamin Bloom (ed.), *Taxonomy of Educational Objectives,* Handbook I: Cognitive Domain (New York: David McKay Co., Inc., 1956); and D. R. Krathwohl, B. S. Bloom, and B. B. Masia, *Taxonomy of Educational Objectives* (New York: David McKay Co., Inc., 1964).

[4] See Burton Weisbrod, *External Benefits of Public Education,* An Economic Analysis (Princeton, N.J.: Industrial Relations Section, Department of Economics, Princeton University, 1964).

[5] There is no empirical verification for this assumption.

[6] For a discussion of their relevance to estimating production functions for industry, see Dennis Aigner and S. F. Chu, "On Estimating the Industry Production Function," *American Economic Review* (September 1968), No. 4, pp. 826-839.

[7] Henri Theil, "Specification Errors and the Estimation of Economic Relationships," *Revere Institute Internationale de Statistique,* Vol. 25 (January 1957), pp. 41-51.

[8] As an illustration, Bowles uses the number of days that the school was in session as a proxy ". . .to represent the general level of community interest in and support of education." *op. cit.,* p. 49. Yet, such an indicator is more likely to be governed by

State mandate than by community educational interests, educational support, and political processes. That is, each State requires a minimum session in order for the school district to qualify for aid. Accordingly, the main variance in the measure is accounted for among States. For the national sample used by Bowles the mean for the "days-in-session" variable was 180 and the standard deviation was only 4.

[9] See S. Bowles and H. M. Levin, "The Determinants of Scholastic Achievement—An Appraisal of Some Recent Evidence," *The Journal of Human Resources,* Vol. III, No. 1 (Winter 1968), pp. 3-24.

[10] See John Kain and Eric Hanushek, "On the Value of Equality of Educational Opportunity as a Guide to Public Policy," Program on Regional and Urban Economics, Discussion Paper No. 36, Harvard University (May 1968).

[11] See S. Bowles and H. M. Levin, *op. cit.,* pp. 6-7.

[12] See Janet Elashoff and R. M. Elashoff, "On Regression Analysis with Missing Data," *Computers, Data Bases, and the Social Sciences,* Ralph Bisco (ed.), John Wiley & Sons, forthcoming.

[13] This has been discussed at length by Bowles and Levin in "The Determinants of Scholastic Achievement," and by the same authors in "More on Multicollinearity and the Effectiveness of Schools," *The Journal of Human Resources* (Summer 1968), pp. 393-400.

[14] This has been attempted in Herbert Kiesling, "Measuring a Local Government Service: A Study of School Districts in New York State," *Review of Economics and Statistics* (August 1967), pp. 356-367. Also see James W. Guthrie, *et al., op. cit.,* pp. 135-144.

[15] See Eric Hanushek, *The Education of Negroes and Whites* (Unpublished Doctoral Dissertation, Massachusetts Institute of Technology, 1968).

[16] Following the capital embodiment approach more strictly, Dennis Dugan has calculated the monetary value of parents' educational investment in their offspring by calculating the opportunity cost or market value of such services. The values of father's educational investment, mother's educational investment, and school investment (all measured in dollars) seem to have high combined predictive value in explaining achievement levels. See Dennis Dugan. "The Impact of Parental and Educational Investments Upon Student Achievement." Paper presented at 129th Annual Meeting of the American Statistical Association (New York City, August 21, 1969), mimeo.

[17] That is, the residual term is likely to be correlated with $A_{2_{it}}$, $A_{3_{it}}$, and $A_{4_{it}}$, and the direct application of the ordinary least-squares estimator will not yield unbiased estimates of the structural parameters of equation (2). See J. Johnston, *Econometric Methods* (New York: McGraw-Hill, 1963), Chapter 9.

[18] A description of the identification problem is found in J. Johnston, *op. cit.,* pp. 240-252. Also see Franklin Fisher, "Generalization of the Rank and Order Conditions for Identifiability," Econometrica, Vol. 27 (1959), pp. 431-447).

[19] $F_{i(t)} = [f_{1_{i(t)}}, f_{2_{i(t)}}, \ldots, f_{n_{i(t)}}]$

That is there are n elements in the $F_{i(t)}$ vector, but not all of them are germane to any particular equation.

[20] The parents' attitude variable might be considered to be an intermediate output in that its social value is more a function of its effectiveness in producing other outputs rather than its use as an end in itself. In a similar vein the teachers' attitudes might be introduced into the model as an endogenous variable.

[21] These data were derived jointly with Stephan Michelson at The Brookings Institution from magnetic tapes provided by Alexander Mood. The same set of data is used in the Michelson paper, contained in this volume.

[22] See Henri Theil, *op. cit.*

[23] For contrasting views on the extent to which innate traits are genetically determined with particular emphasis on "intelligence," see J. McV. Hunt, *Intelligence and Experience* (New York: Ronald Press, 1961); and Arthur R. Jensen, "How Much Can We Boost IQ and Scholastic Achievement?" *Harvard Educational Review,* Vol. 39, No. 1 (1969), pp. 1-23. See also Gerald Lesser and Susan S. Stodolsky, "Learning Patterns in the Disadvantaged," *Harvard Educational Review,* Vol. 37, No. 4 (1967), pp. 546-93.

[24] See Nevin S. Scrimshaw, "Infant Malnutrition and Adult Learning," *Saturday Review,* Vol. 51, No. 11 (March 16, 1968), pp. 64-66.

[25] For more information on this relationship see Henry M. Levin, *Recruiting Teachers* to be published by Charles E. Merrill. Also see Levin, "A Cost-Effectiveness Analysis of Teacher Selection," *The Journal of Human Resources,* Vol. V, No. 1 (Winter 1970), pp. 24-33.

[26] Under certain conditions the simultaneous equation estimates are subject to greater specification biases than the ordinary least-squares ones. See Robert Summers, "A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equations Estimators," *Econometrica* (January 1965), pp. 1-47. Also see Franklin M. Fisher, "The Relative Sensitivity to Specification Error of Different k-Class Estimators," *The Journal of the American Statistical Association,* Vol. 61, No. 314, Part 1 (1966), pp. 345-347. Stephan Michelson has shown results for alternative specifications of the single equation model in *op. cit.,* published in this volume.

[27] See J. Johnson, *op. cit.,* pp. 231-236.

[28] See Patricia Sexton, Feminized Male: Classrooms, White Collars, and the Decline of Manliness (New York: Random House, 1969). As we might expect, females show lower grade aspirations (table 4).

Chapter 4

THE PRODUCTION OF EDUCATION, TEACHER QUALITY, AND EFFICIENCY

Eric Hanushek

It is currently in vogue to claim that the public education system is failing us. This is supported by a variety of evidence on incomes, racial disparities in achievement, and so forth. However, such statements by themselves are not very useful since, even if true, they provide the educational decisionmaker with no information from which to do his job better. It is simply easier to provide a balance sheet of the outputs of education than it is to provide prescriptions for action, and this fact accounts for why there has been more analysis of the results of education than of methods of improving education.

Hopes for improving public education in the United States depend upon our learning from past experiences. We must be able to assimilate the results of past educational programs and past instruction. However, the complexities of education make this assimilation very difficult. School administrators are often good at making judgments about very specific aspects of education. For example, a principal often can make a good judgment about which teachers are getting results and which are not. Yet, at the same time he has difficulty in pinpointing the characteristics which lead to "getting results." He will often conclude that it's all in the individual. But, if this is truly the case, we have little hope for improving public education. In order to improve our educational system we must be able to make some generalizations about characteristics of teachers which are more or less favorable to education.

This paper looks at the educational process with the aim of identifying the role of teachers in education. Moreover, since the implicit model of education used by administrators is know—namely that a teacher's productivity is a function of experience and educational level, it is possible to make some statements about the efficiency of schools in their hiring of teachers. After sketching a general model of the production of education, the paper presents two separate attempts at estimating models of education. The first relies upon the data from *Equality of Educational Opportunity (EEO);*[1] the second uses a new sample collected from a California school system during the summer of 1969. From these analyses it is concluded that: (1) teachers do generally count in education; (2) schools now operate quite inefficiently; and (3) there appears to be considerable latitude for public policy to improve our educational system.

Conceptual Model

It is not possible to look at the role of teachers in education in isolation. Instead, one must consider all of the factors that enter into educational process and how they interact with one another. Thus this study of the effects of teachers on the education of children rightfully starts with a discussion of a larger model of the educational process and the various factors that enter into it. After presenting an abstract model of the educational process, this section considers specific measurement of the various inputs to the educational process and the outputs of the educational process. If one can identify and measure the effects of schools and teachers on the education of individual children, then one can make some statements on how best to organize the school to provide the most educational output.

The basic model of the educational process can be depicted by an equation such as Equation 1.

(1) $A_{it} = f(B_i^{(t)}, P_i^{(t)}, I_i, S_i^{(t)})$ where

A_{it} = vector of educational outputs of the i^{th} student at time t

$B_i^{(t)}$ = vector of family inputs to education of i^{th} student at cumulative time t

$P_i^{(t)}$ = vector of peer influences of i^{th} student cumulative to time t

I_i = vector of innate endowments of i^{th} student

$S_i^{(t)}$ = vector of school inputs to i^{th} student cumulative to time t

This model simply states that educational output (A_{it}), itself a multidimensional factor, is a function of the cumulative background influences of the individual's family ($B_i^{(t)}$), of the cumulative influences of his peers ($P_i^{(t)}$), of his innate abilities (I_i) and of the cumulative school inputs ($S_i^{(t)}$). While this abstract model is not very operational, it does provide a framework for discussion of models of the educational process which can be tested empirically.

Specific measures of each of the inputs listed in Equation 1 are derived from a combination of past work in the field, theoretical considerations, and sheer data availability. For instance, one can think of many measures of the output of the educational process. It would be possible to use standardized test scores, juvenile deliquency rates, future income streams, or level of education completed. However, for any given sample of data one is usually hard pressed to find more than one of these specific measures. While theoretically one thinks of schools producing several different outputs, usually lumped under the major categories of cognitive development and socialization, the availability of data has restricted most past studies to examining a single output. Indeed, this will be the situation in the analysis that is presented in this paper. This paper concentrates entirely on an analysis of cognitive development as reflected in scores on standardized ability and achievement test scores.[2] It is believed that these scores represent differences which are valued by society.[3]

The inputs are subject to many of the same considerations as the measure of output. There is no firm theoretical basis for choosing inputs. Likewise, there is often a lack of desired data. Each input vector will be discussed in turn.

Families contribute to the education of children in many different ways. They provide basic shelter and food for the individual child. But more than that, they provide models of verbal structure, examples of problem solving, and a basic set of attitudes to the individual child. To measure each of these concepts explicitly would be a very difficult task, but for our purposes this is not really necessary. It is widely accepted that the relevant educational inputs are highly correlated with the socioeconomic status (SES) of the family. Thus one can indirectly include the effects of each of these individual family inputs in the educational process by including a set of measures of socioeconomic status. These measures include parents' educations, goods in the home, family size, and father's occupation.

Peer groups provide many of the same inputs that the families provide. The individual child's peer groups would include his friends both inside and outside of school. To be precise, one would want to know exactly which individuals were friends or tended to interact with each other, but collecting this kind of information on a very large scale would be prohibitively expensive. In this case, it seems acceptable to aggregate all classmates of the individual in the classroom or school and take that as the peer groups. In measuring the interactions of individual children one can use the same proxies for peers that are used in the case of the individual's family, that is, use socioeconomic status as a proxy for the types of interaction which exist among friends. Thus for peer groups we would want to take aggregates of the individual family background measures.

Innate ability is probably the most difficult concept to measure in the whole model. In fact, it is not well understood how innate abilities enter into the educational process, and there exists considerable controversy over the role of innate ability in education. The only consensus which appears to exist in the area is that common IQ scores do not do an adequate job of measuring innate abilities. All is not lost, however, when innate abilities cannot be measured directly. In particular, under a set of plausible assumptions (which will be detailed in the empirical section) it is possible to circumvent the most serious problems.

School influences are the focus of this study and will be discussed in more detail than the other inputs. The hypotheses to be analyzed actually are quite simple and straightforward. It is surprising how little is actually known about the ways in which schools and teachers affect education. This largely results from a fixation on inputs to education rather than outputs. However, one can input a set of hypotheses about teacher effects from the behavior of schools. In particular, schools base pay schedules on teaching experience and educational levels. Thus, they must believe that increased experience and further schooling have a positive relationship to educational output. These provide two central hypotheses in the study of the educational process.

Other hypotheses can also be found in the actions of school administrators. A frequent compensatory education plan is the reduction of class size. Since this is a very expensive undertaking, the presumed benefits (increased outputs) must be great. Also there are a large number of people who argue that some forms of student distributions in the schools and classrooms (e.g., ability tracking or racial and social integration) have a beneficial effect on education.[4] All of these are testable hypotheses about the relationship between school inputs and achievement.

Further, in recent literature, particularly *Equality of Educational Opportunity* (EEO), there is a suggestion that one can measure other dimensions of teacher and school quality. These include attitudes of teachers and administrators, verbal facility (and perhaps general ability) of teachers, quality of physical plant, quality of teacher education, background of teachers, and more.

Together, the preceding form the rudiments for a testable model of the educational process. While some modifications are required because of data limitations, this basic structure will hold in the empirical section.

Empirical Analysis

Two separate analyses of the educational process in elementary schools area have been undertaken in this paper. The first relies upon the data for the Northeast and Great Lakes of *Equality of Educational Opportunity*. The second uses a sample drawn from a California school district during 1969. Each of these analyses will be described separately and then they will be compared for consistency and conclusions.

Multisystem School Analysis[5]

The well-known report *Equality of Educational Opportunity* assembled the best data bank on public education to date. This 1965 survey collected a wealth of data pertaining to students, schools, and the outcomes of education. A reanalysis of these data comprises the first section of applications of the basic educational model.[6]

The survey collected data on some 570,000 students and 67,000 teachers across the country. It was a purely cross-sectional survey of students in grades 1, 3, 6, 9, and 12. Minorities were intentionally overrepresented in the sample.

The student information included a set of standardized test scores (verbal ability, nonverbal ability, reading achievement, and mathematics achievements) and questionnaire responses to both objective questions about the students' background and subjective questions about the students' attitudes toward school and society and the parents' attitudes about similar issues.

The teachers in the sampled schools completed a questionnaire concerning objective background characteristics (education, family background, experience, etc.) and subjective characteristics (attitudes toward students, minorities, compensatory education, etc.). They also completed a simple verbal facility test.

Finally, principals and school superintendents supplied information on general school characteristics, curriculums, and their personal backgrounds and attitudes.

In using these data to test the model of the educational process, two factors are immediately evident. The data do not relate school and teacher inputs to individual students. In no place is there any information on specific inputs received by or available to an individual student. One only knows what school averages look like. Therefore, there would be considerable error in the school input variables if one attempted to estimate a model for individuals like Equation 1. Secondly, there is no measure of innate abilities in the model.

The first problem, the inability to estimate models for individual students, is overcome by looking at total school models. Instead of using the achievement of individual students as the output of the educational process, students are aggregated across schools so that average scores for a given grade represent the output. At the same time, inputs are aggregated across the school so that average background characteristics and average school characteristics form the inputs. This tends to minimize the data problems introduced by incompatibility of student and school data.

One obvious loss from this aggregation is the influence of peers on students. It is no longer possible to differentiate between family backgrounds (in aggregated form) and peer influences. (One crude peer effect can be analyzed. This is the effects of one racial group on others. However, this becomes tricky to interpret because of the intertwined and competing hypotheses involved in the racial influence variables.)

Innate abilities are not handled as neatly. There is no direct measure. However, at least for whites, it is reasonable to assume that this factor is fairly well captured in the family background variables. This is the case if innate abilities tend to be hereditary and if social mobility is highly correlated with ability.[7] For blacks, where the parent-to-son correlations of SES are not nearly as pronounced, this logic is more strained.[8] The principal problem arising from lack of measure of initial endowments is biased statistical results. But bias only arises when the excluded variable (innate abilities) is not independent from the included inputs. Thus, even in the black case, severe problems at least at the school level do not arise unless there is a mechanism which leads to the correlation of innate abilities and specific school resources. For the purposes of analyzing school and teacher influences this omission, then, does not seem too damaging. Note, however, that this factor further complicates the family background factors. Those who would attempt to derive policy implications from the background portions of the model are warned again of the extremely complicated nature of that set of inputs.

The specific school analysis undertaken involved estimating separate black and white models. Separate models were estimated for two reasons. First, since many of the inputs—particularly the background factors—are measured by social class proxies, there is no reason to assume that these nominal measures imply the same behavioral content. Secondly, there is no reason to assume that the educational process is the same across racial lines. In fact many people maintain strongly that differences do exist.

The analysis is concentrated upon the sixth grade students in the sample. This choice was the result of two factors. The inability to include historical information due to the cross- sectional survey with little data on the past, indicated that data from earlier schooling with less chance of moves, changes in status, etc., introducing error would be superior. However, there was a trade-off here because the students supplied all of the information on their background (no consultation with parents); going back to the first and third grades would introduce a different type of data error. The desirability of using elementary schools for the analysis is immediately obvious. The generally simpler school organization, the more standardized curriculums, and the more homogeneous size make elementary schools much more attractive for modeling than intermediate or high schools.

The samples used for the analysis included all urban elementary schools from the Northeast and Great Lakes regions of the Equality of Educational Opportunity survey that had at least five white or black sixth graders. This yielded 471 schools with five or more white students and 242 schools with five or more blacks. In both samples the racial mix contains observations across the whole spectrum from less than 5 percent of the opposite race to over 95 percent, although both samples are heavily represented by highly segregated schools.

Results—Multisystem School Analysis

Models of education for whites and blacks were estimated using regression techniques.[9] In both cases a multiplicative (log-log) functional form proved superior to a linear form. Thus, the estimated coefficients can be interpreted as elasticities.[10] Three separate measures of teacher quality proved significant in the models: teacher experience, teacher verbal facility test scores, and the percent of students with a nonwhite teacher during the previous year. The effects of teachers on the production of verbal achievement is presented in table 1 along with the means and standard deviations.

TABLE 1
TEACHER EFFECTS ON VERBAL ACHIEVEMENT, MEANS, AND STANDARD DEVIATIONS

Variable	Elasticity	Mean	Stnd. Dev.
WHITE MODEL			
Teacher experience (years)	.020	11.9	4.6
Teacher test score	.117	24.8	1.4
% students with nonwhite teacher last year	–.024	13.4	16.0
BLACK MODEL			
Teacher experience (years)	.045	11.3	4.0
Teacher test score	.178	24.0	1.8
% students with nonwhite teacher last year	–.026	44.7	19.4

Complete model: Verbal = f(goods in home, father's education, family size, attitudes, central city, racial composition, and teachers)

The complete models are found in the appendix. Since the focus of our attention is on the effects of teachers, only teacher effects are shown in table 1 even though the estimates were derived from a larger model. Suffice it to say here that the background variables appear to do a good job of measuring home and peer influences on education. Further, the estimated effects of teacher inputs seem to be invariant to the precise formulation of background factors and to the inclusion or exclusion of the attitudinal variables.

Since the school influences in the two models appear quite similar, it is possible to discuss both models at the same time. One of the more interesting features of the models is that only one factor which is explicitly purchased by schools affects achievement; this is teacher experience. Further, the small coefficients indicate that experience does not have an overwhelming effect on achievement. The existence of "seniority rights" in school selection suggests an upward bias as school achievement could well influence selection by teachers. However, indirect evidence of the insignificance of direct attitude variables about school selection by the teachers indicates that this variable is chiefly a "pure" experience measure. It is somewhat surprising that the elasticity is constant across the whole range of experience, although tests for differences in different ranges proved insignificant.

The teacher verbal test score represents the best measure of teacher quality contained in the data. This provides a method of making standardized comparisons across teachers but is a still crude measure of teacher quality. It gives some measure of the technical competence of the teaching staff in one particular dimension—verbal ability—and it probably acts as a partial proxy for general intelligence. Nevertheless, there are many other dimensions of teaching, e.g., rapport with the class, empathy, warmth, knowledge of subject matter, which are valuable in teaching but not included in this measure.[11] Given these shortcomings, the magnitude of the effect is significant. The elasticity of .12 (.18) for such a poorly measured indicator of teacher quality provides considerable encouragement in the ability of schools to affect children. Table 2 indicates the small variation in this measure; the standard deviation for whites equals only 1.4 with a mean of 24.8 and a maximum score of 30 with a black sample mean approximately one point less. Nevertheless, there are wide fluctuations of scores even within cities. Within one sampled city, there were differences of 40 percent between the best and worst schools.[12] Switching the teacher staffs would result in a 5 to 7 percent increase in average achievement.

The final teacher quality measure is the percentage of sixth graders who had a nonwhite teacher during the last year. This is interpreted as a measure of part of the teacher quality distribution, i.e., the lower end of the distribution. This interpretation arises from our knowledge of the education provided to blacks. Many studies, including a survey of colleges presented in *Equality of Educational Opportunity,* show the general quality gap between Negroes and whites who go into teaching.[13] This not particularly surprising given that blacks are given inferior elementary and secondary school education and then proceed to segregated colleges which tend to widen the educational gap (by race).[14]

Before discussing the larger implications of these results, it is useful to digress for a moment and discuss some of the school factors which proved insignificant in modeling the educational process. These include teacher degree level, sex, age, teaching certificates, attitudes toward teaching and the students, measures of teacher background, and class size. Certainly, there are considerable measurement errors in each and these errors will affect the significance of the various factors. However, none seems to exert a strong influence on achievement.

A few general conclusions arise from this analysis. First, the general low effect of purchased aspects of teachers (advanced education and experience) indicates that schools are acting inefficiently. Since school systems pay handsome bonuses for these attributes, it is only economical to have people with advanced degrees if they contribute a proportionately higher amount to achievement. This does not appear to be the case.

However, these models do not support the contention that schools do not count. To the contrary, they imply that higher quality teachers do produce higher levels of achievement. Further, given the general problem of measurement errors in the data and the crudeness of the variables, the coefficients tend to be underestimated or biased downward.[15] Looking at table 1, there is also the distinct impression that teacher quality impacts more on blacks than on whites. While differences in the coefficients are small, they are consistent. If in fact this is the case, it indicates that schools can increase educational achievement for whites and blacks by allowing for these differences in the educational process. For example, they would be able to increase black achievement without changing white achievement by shuffling teachers with more experience into predominantly black classrooms (and possibly compensating predominantly white classrooms with more verbal teachers).

It is unreasonable to push these models too hard. They make two essential points. First, teachers do appear to matter. Better teachers (better here in a very limited way) achieve better results. Second, schools appear to be inefficient. They appear to be hiring the wrong things.[16]

Single System, Individual Student Analysis[17]

A similar type of analysis was carried out with a different set of data which allowed a more accurate measure of the teacher inputs received by each child. In particular, individual students were matched with individual teachers. This allowed for an historical element to be introduced by matching with past teachers and alleviated the need to estimate school production functions. Thus, the data came much closer to the conceptual model of Equation 1.

The basic sample of data was drawn from a large school system in California during the summer of 1969. All children in the third grade during the school year 1968-1969 were initially included in the sample. For these 2,445 students, information on family background, scores on the Stanford Achievement Tests, and names of teachers was abstracted from cumulative records. At the same time, all kindergarten through third grade teachers currently in the system were surveyed for information fairly similar to that contained in *Equality of Educational Opportunity.* Information was collected on teacher backgrounds, attitudes, and specific aspects of schooling. An attempt was made to ascertain their use of time, i.e., the division in the classroom between instructional efforts, disciplinary efforts, and administration. Also, a verbal facility test was given each teacher.[18] The sample used for this analysis was developed by applying two criteria to this group of all third graders. First, individuals were eliminated from the sample if data were not available on both their second and third grade teachers. Second, students were eliminated if both first and third grade achievement test scores were not available. When these criteria were applied, a total of 1,061 students was left in the sample.

This sample allows another method of dealing with the problem of initial endowments. In particular, since there is a measure of previous test scores, it is possible to restrict the analysis entirely to one period of schooling by including the previous score for an individual as an input into the process. In this matter all of the level determining aspects of innate abilities can be eliminated. This seems to go a long way toward minimizing any biases arising from this missing information.

Looking at one school district has both advantages and disadvantages. Many hard-to-measure attributes of a school such as curriculum, school organization, community attitudes, etc., are automatically taken care of by looking at one school system. Thus, potential biases from community or system specific variables which cannot be or are not measured are eliminated in such a sample. However, the same arguments can be turned around in the other direction. By looking at only one system it is difficult to make generalizations about behavior in other systems located in different regions and having different types of organization. If specific system attributes are very important, it might not be possible to apply estimated models to other systems. This implies that the previous section's analysis and the analysis in this section are very much complements of each other. Each has weaknesses, but consistency in the different samples would strengthen the results considerably.

Empirical Results

For analytic purposes the sample was divided into subsamples. First, whites and Mexican-Americans (the only minority group represented in the system) were separated. This follows the reasoning given for looking at whites and blacks separately. The nominal values of the proxies for background inputs do not necessarily have the same meaning for the two groups, and there is no reason to insist on the same model of the educational process for both groups. Further, the ethnic samples were divided on occupational grounds—fathers in manual or blue collar occupations and fathers in nonmanual or white collar occupations. This left three samples: white, manual occupation (n = 515); white, nonmanual occupation (n = 323); and Mexican-American, manual occupation (n = 140).[19]

The first step in analyzing the data was to estimate third grade achievement (A_3) models using only the teacher inputs which are purchased by the system to represent school effects. Two linear regression models were estimated (one using first grade achievement as an input, the other not using it). The "pay parameters" of years of teaching experience, possession of a master's degree (=1) or not (=0), and the number of college units beyond the highest degree represented the school inputs in the models. These attributes pertained to the specific second and third grade teachers for each student.

As table 2 and table 3 ably demonstrate, there is a general lack of statistical significance of these factors.[20]

TABLE 2: SIGNIFICANCE OF TEACHER EFFECTS (*Gross output*)

A_3 = f(sex, income, siblings, no. absences, percent Mexican-American, aver. income in school, $EXPER_3$, $MASTER_3$, $UNITS_3$, $EXPER_2$, $MASTER_2$, $UNITS_2$)

t statistics

	White Manual	White Nonmanual	Mex-Amer Manual
$EXPER_3$	.74	2.74	–.04
$MASTER_3$	.89	–2.69	–.47
$UNITS_3$	2.04	.21	1.09
$EXPER_2$	–1.39	–.55	.77
$MASTER_2$	1.45	–.15	–.42
$UNITS_2$	2. 26	2.93	–.34

TABLE 3: SIGNIFICANCE OF TEACHER EFFECTS (*Value added*)

$A_3 = f(*) + A_1$

t statistics

	White Manual	White Nonmanual	Mex-Amer Manual
$EXPER_3$	.56	1.69	–.45
$MASTER_3$	.18	–1.91	.59
$UNITS_3$	.94	1.05	1.77
$EXPER_2$	–.61	.30	1.31
$MASTER_2$	1.94	.60	–.00
$UNITS_2$	.31	-.06	–1.60

Only four of 18 coefficients in the gross output case have significant t values; none in the value added case have significant t values. Further, of the significant coefficients, one has the wrong (unexpected) sign. The other three coefficients apply to the number of units beyond the highest degree and, thus, have no meaning when degree level (MASTER) is not included in the model (or has an insignificant coefficient). The implication is immediately obvious—the things that schools are buying do not appear to be valuable in the educational process.

However, the above results give minimum guidance to an administrator. While they indicate what he should not do they give a very imperfect picture of what he should do. For his purposes we wish to identify what attributes of teachers do seem to count. That is the emphasis of the remainder of this section.

Separate models using different measures of teacher characteristics were again estimated for white, white collar; for white, blue collar; and for Mexican-American, blue collar. The results for these groups were quite different. Teacher effects do not appear to be consistent across the three groups.

White Manual

The white manual occupation model comes closest to the previous school models. Equation 2 displays the model of the production of Stanford Achievement Test (Reading) scores estimated for 515 third graders. Variable definitions, means and standard deviations are found in table 4.

TABLE 4
VARIABLE DEFINITIONS, MEANS, AND STANDARD DEVIATIONS – WHITE MANUAL OCCUPATION MODEL

Variable	Mean	Stnd. Dev.	Definition
A_3	55.74	19.1	Stanford Achievement Test raw score — 3rd grade
F	.50	.5	Sex: = 1 for female = 0 for male
R	.08	.3	Repeat grade: = 1 if a grade was repeated; = 0 otherwise
A_1	35.17	15.1	Stanford Achievement Test raw score - 1st grade
D	17.93	18.8	% of time spent on discipline by 3rd grade teacher
T_3	66.90	15.8	*Quick Word Test* score — 3rd grade teacher
Y_3	1.91	1.6	Years since most recent educational experience — 3rd grade teacher
T_2	68.41	19.0	*Quick Word Test* score — 2nd grade teacher
Y_2	2.64	2.6	Years since most recent educational experience — 2nd grade teacher

Third grade achievement is a function of the starting point (first grade achievement, A_1), sex (F), grade repeats (R), and a set of teacher inputs.

$$\text{(2)} \qquad A_3 = 20.8 + \underset{(2.3)}{2.81F} - \underset{(-2.8)}{6.38R} + \underset{(18.8)}{.79A_1} - \underset{(-2.1)}{.07D} + \underset{(2.4)}{.09T_3} - \underset{(-1.5)}{.57Y_3}$$

$$+ \underset{(1.9)}{.06T_2} - \underset{(-2.9)}{.68Y_2} \qquad R_2 = .51 \qquad SE = 13.5$$

Again, the interest here centers on the teacher inputs. The variable D represents the teacher's estimate of the percentage of classroom time spent on discipline. This gives some idea of the intensity of instruction received by the individual student. As expected, this has a negative impact on achievement; as more time is spent on discipline, less is spent on instruction. This suggests that there are noticeable externalities in the classroom and that efforts to reduce discipline time in the classroom would have positive results on achievement. For example, the principal might assume a very high proportion of discipline chores.

Two characteristics of both the second and third grade teachers were significant. Verbal facility test scores and length of time since most recent educational experience of the teacher proved to be important attributes affecting achievment. The third grade teacher elasticity at the point of means of .11 for T and the second grade elasticity of .07 fall in fine with those from the previous school analysis. It is a little surprising, however, that the elasticities are slightly less here than in the other models. The other teacher variable, Y, indicates that recent educational experiences—either undergraduate or graduate level—are important. Thus, efforts to have teachers return to school during summers seem justified in terms of effects on education. The cumulative effect (master's degree and total units) is not as important as recent involvement.

There are some important policy implications surrounding the verbal test measure of teacher quality. By interchanging teachers at the top and bottom of the verbal ability scale for this system, achievement changes by .2 to .4 grade levels.[21] This seems quite significant at this grade level, particularly if the increasing grade level disparities hypothesized in *Equality of Educational Opportunity* hold true for the individuals in this sample.[22] Thus, teacher distribution can have a significant effect on individual children. Further, since this test has national norms, it is possible to get some idea of how the teachers being hired in this system rate when compared with other college graduates. The mean score of 68 places the teachers in this sample slightly under the median for female college graduates. Thus, this system is not being successful in attracting the best people.

White Nonmanual

The model estimated for the 323 children with white collar backgrounds (Equation 3) did not show the importance of teachers to be as high as in the blue collar white sample. Definitions, means, and standard deviations are found in table 5.

TABLE 5
VARIABLE DEFINITIONS, MEANS, AND STANDARD DEVIATIONS– WHITE NONMANUAL OCCUPATION MODEL

Variable	Mean	Stnd. Dev.	Definition
A_3	64.82	16.8	Stanford Achievement Test raw score - 3rd grade
A_1	42.43	15.8	Stanford Achievement Test raw score - 1st grade
C	.19	.4	Clerical occupation: = 1 if father in clerical job; = 0 otherwise
Y_3	2.02	1.7	Years since most recent educational experience - 3rd grade teacher
S_3	7.85	8.1	Years of experience with this socioeconomic level - 3rd grade teacher
Y_2	1.88	1.7	Years since most recent educational experience - 2nd grade teacher
S_2	7.94	6.1	Years of experience with this socioeconomic level - 2nd grade teacher

Equation 3 indicates that, given the first grade achievement of the student, children with fathers in clerical occupations (C) score lower. Further, the recentness of educational experience (Y) is again a factor along with the amount of experience the teacher has had with this socioeconomic level(S).

$$\begin{array}{lllllllll} (3) & A_3 = 35.9 & + .72A_1 & - 5.1C & - .79Y_3 & + .10S_3 & - .66Y_2 & + .20S_2 \\ & & (-3.0) & (-1.9) & (1.2) & (-1.7) & (1.8) & \end{array}$$

$$R_2 = .52 \qquad SE = 11.8$$

Each of these teacher variables is statistically less significant than the teacher variables in Equation 2. Further, the magnitudes of the coefficients suggest that teachers have less effect on these children. The elasticity at point of means for each of the four teacher variables is less than .025. Thus, changing the input values by any reasonable amount yields a considerably smaller achievement change than was found changing teacher inputs in the sample of children in blue collar families.

Mexican—American Manual

In looking at the 140 Mexican—American children, it was impossible to find any discernible impact of schools. The best model of the educational process for these children, Equation 4, shows that in addition to entering achievement scores (A_1), only sex (F), grade repeated (R), and differences in family background (SS and SK) affect third grade achievement. Variable definitions, means and standard deviations are found in table 6.

TABLE 6
VARIABLE DEFINITIONS, MEANS, AND STANDARD DEVIATIONS—MEXICAN-AMERICAN MANUAL OCCUPATION MODEL

Variable	Mean	Stnd. Dev.	Definition
A_3	47.61	19.4	Stanford Achievement Test raw score - 3rd grade
A_1	28.06	12.5	Stanford Achievement Test raw score - 1st grade
F	.54	.5	Sex: = 1 for female = 0 for male
R	.08	.3	Repeat grade: = 1 if a grade was repeated; = 0 otherwise
SK	.34	.5	Skilled labor: = 1 if skilled occupation; = 0 otherwise
SS	.38	.5	Semiskilled labor: = 1 if semiskilled; = 0 otherwise

$$\text{(4)} \quad A_3 = 14.6 + 97A_1 + 2.84F - 8.92R + 8.22SK + 5.96SS$$
$$\qquad (9.7) \qquad (1.2) \qquad (-2.0) \qquad (2.7) \qquad (2.0)$$
$$R^2 = .51 \qquad SE = 13.8$$

None of the measurable factors used in this analysis concerning teachers impacted on these children, at least in the production of reading achievement. This is a shocking result, and not without its policy implications. The system has not been able to provide the type of instruction necessary for these children. Standard teaching methods do not seem to be appropriate in this case.

Individual Student Models

In developing each of the models a set of variables corresponding to some common hypotheses about the education process was also examined. Consistently, the influence of peers (measured by aggregate characteristics of all third graders in the 25 schools for the sample) was found to be insignificant. Peer influences were measured in a number of specific ways. Occupational distribution was depicted by percentage in nonmanual occupation and average income level; ethnic distribution by percent Mexican-American. Further, ability distribution was considered in terms of average achievement scores in the first grade. For teachers, attitudes about compensatory education and minority students proved insignificant. Teacher age, sex, and undergraduate major also showed no effect. Thus, the models displayed imply a set of other hypotheses which proved insignificant.

In terms of teachers the three models can be rank ordered. Teachers have most effect on white children from blue collar families and least effect on children from Mexican-American families. This is disappointing since Mexican-American children are worst off at the beginning of the process (first grade for this analysis). The idea of schools' equalizing initial deficits of these children is obviously not realized.

For the white population teachers obviously do count. Better teachers imply better results. However, better teachers are not measured in the direction that schools measure them by their pay schedules. Instead they are measured in terms of verbal ability, recentness of education and specific socioeconomic class experience. This implies that schools are being inefficient—for a smaller expenditure on teachers schools could reach the same level of achievement. Moreover, there are gains to be made in the school systems from changing their hiring and pay systems.

Conclusions and Implications

The two separate analyses are complements. Each individual analysis has a set of problems associated with it that tends to dilute the findings. However, taken together each appears to make up for the larger problems of the other. Thus, the sum of the two provides a much more reliable picture of education.

Throughout the analysis there is never much question about the ability to model the general educational process, at least as seen in the elementary school. As an overall view of education the models seem to do quite well. The effects depicted are consistent with *a priori* views; the individual elements are statistically significant; and the general explanatory power of the models seems reasonable.

The strongest conclusion from the models is that school systems now operate quite inefficiently. They are buying the wrong attributes of teachers, i.e., attributes which lead to little or no achievement gains. However, it is more difficult to develop the positive side. There are attributes which appear to be quality related which affect achievement. Yet, they can also be interpreted as proxies for other factors. To the extent that verbal facility is just a proxy for general ability or intelligence, then it is not verbal facility which we want to purchase; it is intelligence. Once a hiring policy for verbal ability was instituted, any relationship between verbal ability and intelligence would tend to disappear or possibly reverse. Thus, these models do not provide a practical guide to the school administrator. They only say that there is something there that is desirable for teachers to have.

It is strange to find strong teacher effects for blacks and not Mexican-Americans. This suggests that it is not just deprivation or a lower educational input from outside the school. The most plausible explanation is found in the language problem. There is no measure of the intensity of Spanish language input for each of the Mexican-American children. This omission could obscure any teacher relationship, especially when measured in terms of English reading ability. However, the insignificant effects of schools on these children make it difficult to argue against community control plans for this community.

A large caveat is needed at this point. The only measure of output used in this paper has been achievement test scores. This seems to be very important in terms of further education as that builds upon this foundation. However, this is probably not the only output in schools. In particular, teachers of Mexican-American children may spend a large proportion of their time on socialization aspects of education, e.g., discussing the American heritage or accepted behavioral patterns. This type of instruction by teachers, although somewhat improbable, could lead to the results of Equation 4.

There seem to be a number of directions in which one could proceed at this point. It is obvious that more information about the different dimensions of teacher quality is needed. One must be able to break down the verbal facility measure used in this paper. At the same time it is necessary to develop a model in terms of attributes which the administrator can purchase. While some analysis, particularly that of Levin, suggests that schools implicitly buy attributes such as teacher verbal facility, buying these through a scale in terms of experience and education cannot help but be inefficient.[23] Further, it is evident through comparing verbal scores for teachers with national norms that present salary schedules do not attract the best college graduates into teaching. However, more information is needed about the supply schedules for specific teacher attributes.

At the same time it appears to be very important to expand the measures of output. Achievement test scores certainly do not reflect all dimensions of educational output. The relationship among different outputs of education is very imperfectly understood at this point.

Finally, it is important to broaden the California type sample. It is necessary to develop refined samples over a wide range of experiences. This includes matching students with specific inputs. It is necessary to look at different grades and different school systems. Further, the necessity of refining our measures of teachers is obvious.

APPENDIX

COMPLETE MULTISYSTEM SCHOOL MODELS (verbal ability) (log-log models)

Variable	WHITE Coefficient (t statistic)	BLACK Coefficient (t statistic)
Central City: = 1 if cc = 0 otherwise	−.025 (−4.1)	−.042 (−2.5)
Goods in home (average number with auto, TV, refrigerator, record player and phone)	.599 (10.4)	.662 (7.9)
Father's education (years)	.133 (4.4)	.022 (.4)
People in Home	−.049 (1.8)	−.177 (−3.0)
% who attended nursery school	.015 (4.0)	
% student out migration during past year	−.005 (−1.8)	
% who wish to finish high school or more	.319 (4.8)	.590 (5.5)
% who feel they don't have much chance for success	−.027 (5.9)	−.028 (−2.3)
Racial concentration: = % Negro if between 45 and 75 percent = 0 otherwise		−.011 (−2.5)
Racial concentration: = % Negro if greater than 75 percent = 0 otherwise	−.036 (−3.3)	−.006 (1.3)
% with nonwhite teacher during the past year	−.024 (−7.1)	−.026 (−1.7)
Average score on teacher verbal test	.117 (2.2)	.178 (2.0)
Average years of teaching experience	.020 (3.2)	.045 (2.6)

Acknowledgment

I am indebted to John Jackson for many helpful suggestions.

Footnotes

[1] James S. Coleman, *et al. Equality of Educational Opportunity* (Washington, D.C.: Government Printing Office, 1966), commonly known as the Coleman Report.

[2] Two different tests are used in the course of the analysis: (1) Educational Testing Service's School and College Ability Test (SCAT) for verbal ability in grade 6; and (2) Stanford Achievement Test for reading in grade 3.

[3] There is scattered evidence on this in W. Lee Hansen, Burton A. Weisbrod, and William J. Scanlon, "Determinants of Earnings of Low Achievers: Does Schooling Really Count, Even for Them?", mimeo, Institute for Research on Poverty, University of Wisconsin, February 1969; Burton A. Weisbrod and Peter Karpoff, "Monetary Returns to College Education, Student Ability and College Quality," *The Review of Economics and Statistics,* November 1968; and Randall D. Weiss, "The Effects of Education on the Earnings of Blacks and Whites," Discussion Paper No. 44, Program on Regional and Urban Economics, Harvard University, April 1969.

[4] Cf. U.S. Commission on Civil Rights, *Racial Isolation in the Public Schools* (Washington, D.C.: U.S. Government Printing Office, 1967), Chapter III.

[5] This section relies heavily on analysis presented in more detail in Eric Hanushek, "The Education of Negroes and Whites" (Unpublished Ph.D. dissertation, Massachusetts Institute of Technology, 1968).

[6] The shortcomings of the analysis in *Equality of Educational Opportunity* which suggest a reanalysis would be valuable are discussed elsewhere. Cf. Eric Hanushek and John Kain, "On the Value of *Equality of Educational Opportunity* as a Guide to Public Policy," Discussion Paper No. 36, Program on Regional and Urban Economics, Harvard University, 1968.

[7] Peter M. Blau and Otis D. Duncan, *The American Occupational Structure* (New York: John Wiley and Sons, 1967).

[8] See *The American Occupational Structure.*

[9] Because of the heteroscedastic efforts introduced by using school observations, weighted regression techniques were used to improve the efficiency of the estimators. See "The Education of Negroes and Whites," appendix A.

[10] An elasticity presents the percentage change in verbal achievement that will result from a 1 percent change in the given input.
Mathematically,

$$\text{elasticity} = \frac{\text{\% change in verbal score}}{\text{\% change in input value.}}$$

[11] The narrowness of this quality measure is further attested to by similar analysis of the production of mathematics achievement test scores. In those models the elasticity drops to .09 and the t-ratio goes to 1.3. This indicates a more narrow technical competence interpretation.

[12] The other teacher variables in these schools were roughly equal.

[13] *EEO,* Chapter IV and James A. Davis, *Undergraduate Career Decisions* (Chicago: Aldine Publishing Co., 1965).

[14] *EEO,* Table 3.121.1.

[15] See J. Johnston, *Econometric Methods* (New York: McGraw-Hill Book Co., 1963), pp. 148-150.

[16] This should be qualified somewhat. Even with fixed salary schedules, Henry Levin in *Recruiting Teachers for Large City Schools* (forthcoming) shows that it is possible to estimate supply functions for other characteristics—primarily things like teacher verbal test scores.

[17] The analysis presented in this section is part of an ongoing study of education sponsored by The RAND Corporation. However, this should not be taken to represent the official views of The RAND Corporation.

[18] Edgar F. Borgatta and Raymond J. Corsini, *Quick Work Test: Level 2* (New York: Harcourt, Brace and World, Inc., 1964). This test appears to be superior to the test in *Equality of Educational Opportunity* as it appears to give better discrimination among teachers. One complaint voiced about the *EEO* test is that it was too easy.

[19] These samples are not exhaustive. Children with only mothers or no occupation reported for fathers were not included. For whites, these groups totaled 36 students; for Mexican-Americans, these groups plus the nonmanual occupation group totaled 47. These samples were too small to study separately, and, thus, they were ignored.

[20] When .t. $<$ 1.96, it is not possible to reject the hypothesis that the coefficient equals zero at the 5 percent level.

[21] This is calculated by changing only the third grade teacher verbal score for the lower limit and both second and third for the upper limit. The scores are changed from 40 to 96 to represent the range found in the data. (Maximum score is 100.) The resulting achievement score is then converted to grade level equivalents.

[22] *EEO,* Chapter 3.

[23] See *Recruiting Teachers.*

Chapter 5

TEACHER ATTRIBUTES AND SCHOOL ACHIEVEMENT

George W. Mayeske

In the fall of 1965, at the direction of the 1964 Civil Rights Act, the U.S. Office of Education conducted the most comprehensive educational survey in the history of the American public school system. The intent of the survey was to ascertain whether various racial and ethnic groups have equal educational opportunities.

The survey team collected a comprehensive body of data on public schools and their students, and tried to ascertain the relative importance of different classes of school resources on student achievement. The report of that survey—*Equality of Educational Opportunity*—was issued in the fall of 1966 under the principal authorship of James S. Coleman (Coleman, et. al., 1966).

Since that time, a small staff has been at work at the Office of Education conducting a thorough reanalysis of that same body of data. This paper is excerpted from a larger report of part of that reanalysis, entitled *A Study of Our Nation's Schools* (Mayeske, et. al., 1969).

Several important factors relating to this study should be pointed out at the outset, to be explained in more technical detail later in the paper and, of course, in the full report as well.

First, this study examined a very comprehensive body of data—i.e., the data already collected for the Equality of Educational Opportunity (EEO) survey.

Second, this study had the advantage of considerably more time. Whereas Coleman originally had only about 6 weeks for his analysis, this analysis was conducted over a 3-year period.

Third, this study reduced and combined the more than 400 variables considered in the EEO study to a more manageable number of between 60 and 70 variables. These items were then divided into three main groups: (1) student social background; (2) school characteristics; and (3) school outcomes.

Fourth, this study employed a new technique—the "Commonality Model"—for analyzing the data. The results demonstrated that in analyzing student achievement, very little of the influence of student social background can be separated from their schools. Conversely, very little of the influence of the schools can be separated from the social background of their students. That is, taken in and of themselves, neither student background *nor* school setting can be shown uniquely to contribute a sizeable influence on student achievement. By demonstrating the relationships (or commonality) between the two, however, a high degree of correlation can be shown with achievement.

In conclusion, it may be stated that the overwhelming impression received from these data is that schools are indeed important. It is equally clear, however, that their influence is bound up with that of the student's social background. In such a situation, survey research is of only limited use. More experimental studies are needed, especially of educational innovations. Among such innovations should be included the periodic monitoring of the performance of these programs; the establishment of explicit performance criteria for all school programs; and the establishment of educational institutions that are more balanced in the socioeconomic and racial-ethnic composition of their students.

The Data Base and Background Work for the School Study

The Educational Opportunities Survey entailed the testing and surveying of about 650,000 students in some 4,000 public schools throughout the country in grades 1, 3, 6, 9, and 12, together with their teachers, principals, and superintendents. The data base is comprehensive. Detailed factual and attitudinal information was collected on the students' home background, attitude towards school, race relations, and the world. A battery of ability and achievement tests was administered at each grade level. Information was collected from some 60,000 teachers and 4,000 principals concerning their training and experience, their view of the school, etc. The final part of the teacher questionnaire consisted of a 30-item contextual vocabulary test which was intended to be a measure of the verbal facility of the teacher. In addition, the principal provided data on the school's facilities, staff, programs, curriculums, etc.

The main goal of our background work was to reduce the more than 400 variables in an empirically meaningful way into indexes and sets of indexes. Thus the volume of data processing and complexity of later analyses could be lessened. Before the variables could be reduced into meaningful groupings, however, decisions had to be made concerning the estimation of missing data and the coding or scaling of variables. As a guide in the estimation of missing data or handling of nonresponses, it was decided to analyze the responses to each question against one or more criteria or dependent variables so that not only the percent responding to each item or response alternative, but also their mean score on the dependent variable could be used as a guide in coding the variables and in assigning a value to the nonrespondents. Since the approach differed somewhat for the student, teacher and principal questionnaires in each analysis will be described separately.

Student Analysis

A factor analysis of the five ninth grade achievement measures[1] showed that a single factor could be used to describe their intercorrelations.[2] Accordingly, the weights from the first principal component of the intercorrelations were used to weight scores on the individual tests and sum them to obtain an overall achievement composite. It was this composite which was used as a criterion against which item responses were analyzed. This composite is also the dependent variable for many later analyses.

In order to maximize the linear relationship of each student variable with student achievement, criterion scaling (Beaton, 1969) was employed. In criterion scaling each item response is coded or scaled by assigning the mean value of the dependent variable for each of the different response alternatives for an item.[3]

Teacher Analysis

For the teacher variables, each item was analyzed against the teacher's total score on a self-administered contextual vocabulary test.

Principal Analysis

For the principal's variables, each item was analyzed against the number of students enrolled in the school, the rural-urban and socioeconomic status of the school, and the principal's salary. These analyses were used as guides in assigning codes or scale values and in estimating missing data.[4]

Intercorrelations

First, intercorrelations were established. Then to obtain meaningful groupings of variables, the intercorrelations of the student, teacher, and principal sets of variables were each subjected to a series of factor analyses. The Principal Component technique was used to extract components, and the Varimax technique was used to rotate components having a root of one or greater (Horst, 1965). This approach was essentially iterative; that is, variables that did not form meaningful groupings or blurred an otherwise meaningful grouping were eliminated and the remaining variables refactored. The teacher and student variables readily fell into meaningful groupings after two iterations which resulted in the elimination of about six to 12 variables from each set. The highest weights from the Varimax rotation were used to combine the variables to obtain index scores. In order to keep the index score intercorrelations low a variable was allowed to have a weight on only one index.

The variables from the principal's questionnaire dealt with a wide variety of different aspects of the school. These variables did not readily fall into any naturally meaningful groups. Consequently, *a priori* groupings, such as variables concerned with the physical plant or instructional facilities were subjected to a Principal Component analysis. The weights from the first Principal Component were then used to obtain index scores for each school.

A brief description of the indices obtained and other variables retained for future analyses are given in the appendix. The "full set of school variables" referred to below means the combined set of 31 teacher, principal, and school indexes that are given in the Appendix. Using these indexes we have conducted extensive among-school analyses, i.e., analyses of average difference among schools rather than within each school. These analyses used ninth grades only as the unit of analysis. Thus, in this paper:

- "Socioeconomic Status" refers to the *average* of the Socioeconomic index scores for ninth grade students in a particular school;
- "Achievement" means the *average* achievement of ninth grade students in a particular school;
- "Experience" or "Training" is the *average* Experience or Training of teachers appropriate for students in that school and grade level.

There were 923 schools and 133,136 students used in these analyses.

The Commonality Mode.

Having thus reduced and combined the number of variables the next step was the development of an analytic model. At about the time we were beginning the School study, Alexander Mood developed a technique for the partition of multiple correlation which was to have profound implications for our work. This technique, which we were to discover had been developed independently by Newton and Spurrell (1967), may be described as follows:

> Suppose we have a set of student body variables, B, and a set of school variables, S, and we want to ascertain the contribution that the S variables make to student Achievement after adjusting Achievement for differences in the B variables. Upon performing this operation in the reverse order we find that the contribution of the S variables is small. However, performing the operation in the reverse order we find that the contribution of the B variables is small. We say that the contribution is small in that the squared multiple correlation for each set of variables is large. (Squared multiple correlation refers to the Achievement accounted for by a particular set of variables.) We conclude, therefore, that there must be a high degree of overlap in the way these sets of variables relate to Achievement. To express this quantitatively:
>
> Let: $C(B,S)$ stand for commonality or overlap of the student body variables (B) and school variables (S) as they relate to Achievement
>
> $R^2(B)$—the squared multiple correlation of the student body variables with Achievement
>
> $R^2(S)$—the squared multiple correlation of the school variables with Achievement
>
> $R^2(B,S)$—the squared multiple correlation of the student body and school variables with Achievement
>
> $U(B) = R^2(B,S) - R^2(S)$, that portion of the squared multiple correlation uniquely attributed to the student body variables
>
> $U(S) = R^2(B,S) - R^2(B)$, that portion uniquely attributed to the school variables
>
> then $C(B,S) = R^2(B,S) - U(B) - U(S)$[5] and $R^2(S)$ and
>
> $R^2(B)$ can be expressed as
>
> $R^2(S) = C(B,S)\ U(S)$
>
> $R^2(B) = C(B,S)\ U(B)$

In the following pages these results are "unitized" by dividing the unique and common portions by the squared multiple correlation obtained for both sets of variables combined (viz.

R^2(B,S)). This "unitizing" operation converts the unique and common portions so that they sum to 100 percent.

In its strictest sense this common portion represents an indeterminate situation. That is to say, we cannot tell to which of the two sets, B or S, all or some part of this common portion should be attributed.

The School Study

The objective of the full study (Mayeske, et al., 1969) was to determine those aspects of schools which might be most effective in promoting not only student achievement but also student motivation. However, this paper focuses only on the results for Achievement.

We found that 36 percent of the differences among students in their Achievement is associated with the schools they attend.[6] This leaves 64 percent to be explained by within school and nonschool factors. In the analyses that follow, the 36 percent will be the base or the maximum amount that can be explained.[7] That is, if we were to obtain a multiple correlation of one between student body and school factors and Achievement then we would have explained the entire 36 percent.

Part of the attempt to ascertain the influence of school variables on Achievement was to take into account the kinds of students that the schools get initially. For example, if school "X" had children from families where intellectual activities were not valued or pursued and school "Z" had children from families where these activities were valued or pursued, then one would expect the students in school "Z" to have higher Achievement levels than students in school "X." These differences could be attributed to the influence of the different families rather than to the schools. Thus, schools were equated for differences in the family Social Background of their students prior to looking at the possible influence of school variables on Achievement. The indexes of Socioeconomic Status, Family Structure and Stability, and Racial Ethnic Group Membership were used to represent the Social Background of students. Hereafter, these indices will be referred to as the set of Student Body Social Background (B) variables. Possible school influences include the comprehensive set of 31 school variables given in the Appendix. This set will hereafter be referred to as the School set (S).

As described in the development of the Commonality Model, when the B and S sets were entered into the regression, large squared multiple correlations were observed for each set alone as well as in combination. The portion of variance that could be uniquely associated with one or the other set, however, was small relative to the magnitude of these correlations. This suggested that there was a high degree of overlap or confounding in the way these two sets of variables related to the dependent variable. To express this overlap we performed a commonality analysis for which the "unitized" results are given in table 1.[8] In this table the U(Xi) denotes that portion of the "explained" variance (viz. $R^2(B,S)$) that has been uniquely attributed to the B or S set, while C(B,S) indicates the portion that is in common. The unique portion for one set, say B, and the common portion sum to the percent of explained variance accounted for by that set (e.g. 12 plus 82 or 94 is the portion of explained variance accounted for by B). Similarly, the two unique portions and the common portion sum to 100. All values have been rounded to two places of decimals and leading decimal points omitted.

The really outstanding aspect of the results in this table is the large percentage of overlap or confounding that exists among the B and S variables. We can't really tell to which one of the sets this value of 82, or some part of it should be attributed. The other values are much smaller in magnitude with the unique portion for the S set being 6 percent and for the B set, 12 percent. Using this kind of analysis, one can only conclude that most of the influence of the schools is bound up with the Social Backgrounds of their students and vice-versa.

To further illustrate this latter point we can observe the role that Other School Outcomes (0) play in conjunction with the B and S sets. By Other School Outcomes we will mean the four attitudinal and motivational indexes of: Expectations for Excellence; Attitude Toward Life; Educational Plans and Desires; and Study Habits (see appendix). Results of commonality analyses using these three sets of variables are given in table 2. For three sets of variables there will be a unique value for each set, a value for each of the pairwise combinations (viz. B and S, B and O, and S and O) and a value for the three-way combination (BSO).

Inspection of table 2 shows again that most of the variance in Achievement explainable from the B, S, and O sets is confounded. The portions uniquely attributable to B, S, and O are 7, 3, and 2 percent respectively. That leaves 88 percent (100 minus 7 plus 3 plus 2) as being involved in the higher order combinations. For the two way combinations a large amount (30 percent) is involved in B and S, with 5 and 2 percent for the BO and SO combinations. Just over half of this explained variance is in the *three way combination* of B, S and O. From these observations we can conclude that most of the influence of the schools on Achievement is bound up with the Social Background and motivational levels of the students they get initially (and vice-versa).

We might ask then if there is some subset of S for which this overlap or confounding is greatest. Perhaps this would give us a rough idea of those aspects of the schools that are wielding the greatest influence. Table 3 gives the results of commonality analyses for four sets of variables where the S set has been broken down into the three subsets of School Personnel (T), Pupil Programs and Policies (P), and Plant and Physical Facilities (F). The indexes comprising each set are given in the appendix. As with earlier analyses, there is a value for each higher order combination.

Inspection of table 3 shows that the areas of overlap are greatest when the B and T sets are involved and negligible elsewhere. The largest value (56 percent) is for the two way combination of B and T. The other two way combinations are small to negligible. The three way combinations of BTP and BTF also show moderate values as does the four way combination BTPF. Table 3 shows clearly that the sets for which the confounding is greatest are those where B, the Student Body Social Background, and T, the School Personnel, are present. The Pupil Programs and Policies (P) and Facilities (F) sets show moderate values only in conjunction with the B and T sets.

We might ask then if there are any particular aspects of the School Personnel (T) set for which this confounding is greater, Table 4 gives commonality analyses of the B and S sets with Achievement when the Racial-Ethnic Composition of the teaching staff is deleted from the S set.

When the results in table 4 are compared with those in table 1 we note that the coefficient of overlap drops by 14 percent, the unique portion for B increases by 15 percent and the unique portion for S decreases by 1 percent. What was at first attributed to overlap or confounding has now become attributed to the Student Body Social Background (B). Other analyses showed that as we eliminated "social condition" type variables from the S set—such as Free Lunch and Milk Programs, and the index called Teaching Conditions (i.e. the teacher's view of how much effort the students put forth to achieve, how readily they can maintain order, the extent of student disciplinary problems, etc.)—the coefficient of overlap as well as the unique portion for S tended to decrease while that for B tended to increase.

Still other analyses showed that after schools were equated for their student's Social Background, other variables continued to have relationships with Achievement. These variables were: verbal skills of the teaching staff; teachers' annual salary level; teachers' racial-ethnic composition; teaching conditions; and special staff and services. Although these relationships were not large they were suggestive. However, some of the variables were shown to be closely related to each other. When some of the possible determinants of individual teacher's verbal skills were examined, for example, it was found that their racial-ethnic group membership accounted for a very large portion of these verbal skill differences. Indeed, the existence of a dominant color-caste system in the preparation of teachers was discovered and the self-perpetuating role that it could play through the reinforcement of differential verbal skills along racial and ethnic lines was suggested whereby teachers tend to teach students from the same socioeconomic and racial-ethnic background as their own.

An Interpretation of the Measure of Confounding

We have seen that a large degree of overlap or confounding exists between a school's resources and a student's Social Background as they relate to Achievement. It is suggested that part of this confounding reflects the nature of the educational process whereby students from the higher socioeconomic strata who have an intact family structure and happen to be white or Oriental enter school with more fully developed skills and motivation which enable them to benefit more from their schooling than their less privileged counterparts. Support for this line of reasoning comes from some of our own analyses utilizing the time dependent aspects of the EOS data as well as work by Shaycoft (1967).

Using the time dependent aspects of the EOS data[9] it was found that after schools were equated for differences in the Achievement levels of their first grade students, the measure of confounding or overlap between B and S was larger than their unique portions at the third grade. By the sixth grade, although the unique portions of B and S increased very little their common portion almost doubled its value from what it was at the third grade. Another way of saying this is that the longer the students are in school, even though they start out at the same level of Achievement, the larger becomes the coefficient of overlap or confounding between the B and S sets. A study by Shaycoft (1967) using data taken from the same students measured at two points in time tends also to support the results obtained in these analyses. Shaycoft found that after equating or equalizing students for their initial achievement, students from the higher socioeconomic strata showed greater gains on a later testing than did students from the lower socioeconomic strata.

What we are suggesting, then, is that this measure of overlap represents mainly the interaction of the student's Social Background with the school's staff and, to a lesser extent, also with the school programs. We cannot be more precise about what part of this overlap is due to this kind of interaction for there are also other factors at work. For example, we find even at the first grade that relationships exist between the Achievement levels of the entering students and the attributes of the schools they attend. Thus, schools with entering students of higher Achievement levels have associated with them teachers with higher verbal skills who tend to be white and express a preference for working with high ability students, etc.

We find further that these teacher relationships with Achievement tend to increase at the higher grade levels. Similarly, the relationships of students' Social Background with Achievement increases at the higher grades. This phenomenon suggests what we would like to call the *"ecological-functional dilemma"* in studying school influences. At the beginning of the first grade, students are allocated into schools on the basis of their Social Background. Certain relationships are observed between the attributes of the students and their schools. This we call and *ecological* relationship. Over time, since students with a higher Social Background benefit more from their schooling, ecology and the school's influences (or what we have chosen to call functionality) become more and more interwined so that it becomes increasingly more difficult to separate out their independent influences.

Do Schools Have Important Influences on Their Students?

What these analyses have shown, we believe, is that the schools reflect a deep seated social problem which permeates almost every aspect of our society. This problem, in the main, is that a child's birth into a particular stratum of our social structure largely determines where he will and will not go in the scheme of things. The problem is made even more difficult, however, because one's skin color and language habits tend to be associated with one's position in this social structure.[10] If this interpretation has any validity then it does not seem likely that the schools alone can rectify the problem although they may play an ameliorative role. It seems more likely that the problem warrants a concerted attack from many different sectors of society (viz. jobs, housing, schooling, etc.).

Given that a concerted effort is warranted we might ask what role the schools can play in this effort. We have seen that as the schools are currently constituted very little of their influence can be separated from the Social Background of their students and very little of the Social Background of students can be separated from the influence of their schools. This should not be construed to mean that schools do nothing for their students. Schools do a great deal for all students and this was dramatized in a recent study of children in Prince Edward County, Va. (Green, 1964) who had their schooling interrupted for a few years. When the test performance of these children was compared with children of a comparable background (low socioeconomic status) in a neighboring county it was found that they were 16 to 30 points lower on an IQ test, which was used as a measure of learning. In addition, the young children who would have ordinarily completed the first few grades but who had been unable even to start school, could not even hold a pencil nor follow directions, let alone take a test. Thus schools, even in conditions of poverty, do have important influences on their students. The problem is how to increase the influence of the schools to overcome the effects of these social background barriers.

When we focus on those changes in the schools which have resulted in some degree of success (e.g. language enrichment, remedial reading) we find that these changes were usually on a limited scale and are difficult to repeat even in similar settings (Hawkridge, et al., 1968). These experiences, coupled with the observation that the influence of the schools that is independent of student Social Background is very small, suggest that we should be trying new approaches that differ radically from past practices in situations so structured that the results of the innovations can be clearly ascertained. A range of innovations has been proposed including greater socioeconomic and racial balance among student bodies and teaching staffs; intensified further training of teachers of the disadvantaged perhaps coupled with pay supplements; schools that focus mainly on reading and mathematics; boarding schools; and competitive schools or some form of voucher system whereby the student and his family can select services from a variety of sources. These are all ideas worthy of trial. Some may fail, but the greatest failure of all is not to try, for no one currently knows the magnitude of the role schools can play in helping to ameliorate this deep seated social problem.

Table 1.–Unitized Commonality Analyses of B and S Variables With Achievement

	B	S
U(Xi)	12	6
C(BS)	82	82

$R^2(BS) = 87$

Table 2.–Unitized Commonality Analyses of B, S, and O Variables with Achievement

	B	S	O
U(Xi)	7	3	2
C(BS)	30	30	
C(BO)	5		5
C(SO)		2	2
C(BSO)	51	51	51

$R^2(BSO) = 88$
$R^2(BS) = 87$

Table 3.–Unitized Commonality Analyses of B, T, P, and F Variables With Achievement

	B	T	P	F
U(Xi)	12	2	1	0
C(BT)	56	56		
C(BP)	2		2	
C(BF)	0			0
C(TP)		1	1	
C(TF)		0		0
C(PF)			1	1
C(BTP)	14	14	14	
C(BTF)	4	4		4
C(BPF)	1		1	1
C(BTPF)	6	6	6	6

$R^2\ (BTPF) = 87$

Table 4.–Unitized Commonality Analyses of B and S With Achievement When the Racial-Ethnic Composition of the Teaching Staff is Deleted From S

	B	S
U(Xi)	27	5
C(BS)	68	68

$R^2(BS) = 86$

APPENDIX

Student Indexes

1. Expectations for Excellence—student believes that his mother, father, and teacher want him to be a good student and he desires to be a good student;
2. Socioeconomic Status—defined by mother's and father's educational level, father's occupational level, rooms in the home, number of siblings, reading materials, and appliances in the home, and urbanness of background;
3. Attitude Towards Life—a student with a high score on this index believes that people like himself have a chance to be successful, when he tries to get ahead he won't experience many obstacles, hard work is more important than good luck for success, won't have a hard time getting a job with a good education, etc.;
4. Family Structure and Stability—a student with a high score has both his father and mother in the home, father is the major source of income, he hasn't changed schools recently, etc.;
5. Educational Desires and Plans—a student with a high score desires and plans to go to college, his parents want him to go to college, and he has high occupational level aspirations;
6. Study Habits—a student with a high score spends about 2 hours a day studying, has frequent discussions about his school work with his parents, was read to as a child before he started school, read many books during the summer, etc.;
7. Racial-Ethnic Differences in Achievement—a variable created by assigning each student the average achievement score obtained by his racial or ethnic group.

Teacher Indexes

1. Experience—comprised of the teacher's age, years of teaching experience, and years of teaching in his present school;
2. Teaching Conditions—comprised of various aspects of the teacher's view of his teaching situation such as how hard

the students try to achieve, their academic ability, the reputation of the school, and student disciplinary, racial problems, etc.;

3. Localism of Background—a teacher with a high score has spent most of his life in a small geographic area and has graduated from high school and college in that locale;
4. Socioeconomic Background—comprised of the teacher's parent's educational level, father's occupation and, rural-urbanness of their background;
5. Training—comprised of the teacher's highest degree held, certification, salary level, and tenure;
6. College Attended—comprised of the kind of undergraduate institution attended (e.g. normal school, public or private university, etc.) the highest degree offered by that institution, and teacher's rating of the academic level of the institution;
7. Teaching Related Activities—comprised of the hours of unofficial time spent in preparation for class and counseling, the number of educational journals read regularly, etc.;
8. Preference for High Ability Students—teacher prefers to work with students of higher ability, socioeconomic status, etc.;
9. Sex—scored high for a female, low for a male;
10. Racial-Ethnic Differences in Contextual Vocabulary—a variable created by assigning each teacher the average vocabulary score obtained by his racial or ethnic group;
11. Vocabulary Score—total number of items correct.

Principal and School Indexes

1. Principal's Experience—comprised of age, number of years experience as a principal, etc.;
2. Principal's Training—comprised of the highest degree held and salary level;

3. Principal's College Attended—same as teachers index;
4. Principal's Sex—a variable scored high for female, low for a male;
5. Plant and Physical Facilities—area of plant, possession of auditorium, gymnasium, etc.;
6. Instructional Facilities--special labs, shops, volumes in the library, etc.;
7. Specialized Staff and Services—art, music, and remedial reading teachers, etc.;
8. Tracking—use of various kinds of ability grouping techniques;
9. Testing—frequency of different kinds of testing;
10. Transfers—number of students transferring in and out;
11. Remedial Programs—percent of students in remedial math and reading;
12. Free Milk and Lunch Programs—percent of students who get free milk and lunch;
13. Accreditation—whether or not school has State and regional accreditation;
14. Age of Texts—age of different texts used;
15. Availability of Texts;
16. Age of Building—a variable;
17. Pupils per room—a variable;
18. Pupils per teacher—a variable;
19. Number of students enrolled in the school;
20. School Reputation—the principal's estimate of the school's reputation.

Definition of Sets of Variables

School (S)—11 Teacher indexes plus 20 Principal and School indexes—31 variables

Plant and Facilities (F)—Principal and School indexes 5, 6, 16, and 17—4 variables

School Personnel (T)—the 11 Teacher indexes plus Principal and School indexes 1, 2, 3, 4, 7, and 20—17 variables.

Pupil Programs and Policies (P)—the 10 Principal and School indexes not included in F and T items—10 variables

Student Body Social Background (B)—Student indexes 2, 4, and 7—3 variables

Other School Outcomes (O)—Student indexes not included in (B) above —4 variables.

Development of Measures of Commonality for Three Sets of Variables

Consider the case where there are three sets of variables: a set of Student Body Background variables (B); a set of School variables (S) and; a set of other Outcome measures (O). Then the first order commonality coefficient or portion of the squared multiple correlation that is uniquely associated with a given dependent variable is:

$$U(B) = R^2 (B,S,O) - R^2 (S,O)$$

$$U(S) = R^2 (B,S,O) - R^2 (B,O)$$

$$U(O) = R^2 (B,S,O) - R^2 (B,S)$$

where $R^2(\ )$ represents the squared multiple correlation for the particular set of variables in parentheses with the dependent variable.

The second order commonality coefficients are given by:

$$C(BS) - R^2 (B,S,O) - R^2 (O) - U(B) - U(S)$$

$$C(BO) - R^2 (B,S,O) - R^2 (S) - U(B) - U(O)$$

$$C(SO) - R^2 (B,S,O) - R^2 (B) - U(S) - U(O)$$

and the third order commonality coefficient of which there is only one, is given by:

$$C(BSO) = R^2(B,S,O) - R^2(B,S) - R^2(B,O) - R^2(S,O) - U(B) - U(S) - U(O)$$

The squared multiple correlation for any single set can then be expressed as a function of its different order commonality coefficients. For example, the squared multiple correlation for the Outcome set ($R^2(O)$) can be expressed as:

$$R^2(O) = C(BSO) + C(BO) + C(SO) + U(O)$$

Development of Measures of Commonality for Four Sets of Variables.

Let the four sets of variables be denoted by X_1, X_2, X_3, and X_4. Then the unique portion of first order commonality coefficients for the i^{th} set is given by

$$U(X_1) = R^2(X_1 X_2 X_3 X_4) - R^2(X_j X_k X_l)$$

where $R^2(\ )$ represents the squared multiple correlation for the particular set of variables in parentheses with the dependent variable. As an example, the unique portion for the fourt set would be written as

$$U(X_4) = R^2(X_1 X_2 X_3 X_4) - R^2(X_1 X_2 X_3)$$

There is one unique value for each set of variables, namely four in this case.

The second order commonality coefficient is given by

$$C(X_i X_j) = R^2(X_1 X_2 X_3 X_4) - R^2(X_k X_l) - U(X_i) - U(X_j)$$

As an example, the second order commonality coefficient for the third and fourth sets is

$$C(X_3 X_4) = R^2(X_1 X_2 X_3 X_4) - R^2(X_1 X_2) - U(X_3) - U(X_4)$$

There is one second order commonality coefficient for each combination of sets, namely six in this case.

The third order commonality coefficient is given by:

$$C(X_i X_j X_k) = R^2(X_1 X_2 X_3 X_4) - R^2(X_1) -$$

$$C(X_i X_j) - C(X_i X_k) - C(X_j X_k) - U(X_i) - U(X_j) - U(X_k)$$

There is one-third order commonality coefficient for each three way combination, namely four in this case.

The fourth order commonality coefficient, of which there is only one, is given by:

$$C(X_1 X_2 X_3 X_4) = R^2(X_1 X_2 X_3 X_4) - R^2(X_1 X_2 X_3)$$

$$R^2(X_1 X_2 X_4) - R^2(X_1 X_3 X_4) - R^2(X_2 X_3 X_4) - R^2(X_1 X_2) -$$

$$R^2(X_1 X_3) - R^2(X_1 X_4) - R^2(X_2 X_3) - R^2(X_2 X_4) - R^2(X_3 X_4) -$$

$$U(X_1) - U(X_2) - U(X_3) - U(X_4)$$

The fourth order coefficient can be verbally described as the squared mutliple correlation for all four sets $R^2(X_1 X_2 X_3 X_4)$ minus the sum of the four third order commonalities $C(X_j X_k X_l)$, minus the sum of the six second order commonalities $C(X_j X_k)$, minus the sum of the four unique portions.

Consequently, the squared multiple correlation for the X_4 set can be represented as the sum of its unique value and its different order commonalities, thus:

$$R^2(X_4) = C(X_1 X_2 X_3 X_4) + C(X_1 X_2 X_3 X_4) + C(X_1 X_3 X_4) +$$

$$C(X_2 X_3 X_4) + C(X_1 X_4) + C(X_2 X_4) + C(X_3 X_4) + U(X_4)$$

Computational Formula for the Percent of Variance Associated With the Schools Students Attend

The correction for the appropriate degrees of freedom is a modification of the shrinkage formula for a multiple correlation. (See Thorndike, 1949, p. 204). To use this formula each school is regarded as a dummy variable or pseudo variable where a student is assigned a 1 if he attends that school and a O otherwise. This results in one dummy variable for each school and the dependent variable is regressed against the dummy variables. The formula used is:

$$P^2 = 1 - \frac{(N-1)(1-R^2)}{N-P} \quad \text{where}$$

P^2 = the corrected squared multiple correlation
N = the number of students
n = the number of schools
$P = n - 1$
R^2 = the ratio of the among school variance (S^2_A), to the total variance (S^2T); $S^2_A / S^2T = R^2$

FOOTNOTES

[1] The tests were: General Information; Reading Comprehension; Mathematics Achievement; Verbal Ability; and Nonverbal Ability.

[2] The first principal component of the intercorrelations accounted for 75 percent of the variance.

[3] Almost all of the ninth grade student variables were coded in this manner. When the results of this scaling technique were compared with a more conventional procedure it was found that they were very similar except for some of the attitudinal items which were linearized by the criterion scaling procedure.

[4] However, for the teachers' and principals' questionnaires the items were not coded so as to maximimize their relationship with their dependent or criterion variables.

[5] A generalization to three and four sets of variables is given in the Appendix.

[6] See the Appendix for the specific computational formula used to obtain this value.

[7] If there were no differences among schools this percent would be zero. This would be true whether schools were equally good, equally bad, or equally mediocre.

[8] Results identical to these were obtained when a more conventional coding technique was used for the student questionnaire items.

[9] These were schools for which data was available on their first, third, and sixth grade students. The first grade students were considered as a surrogate for what the third and sixth grade students were like when they entered first grade. The third grade students were considered as a surrogate for what the sixth grade students were like when they were in the third grade, etc.

[10] Although as two large-scale studies have shown (Husén, Plowden), where skin color is not an issue social class is still very much an issue in the benefits students accrue from their schooling.

References

Beaton, A. E. Criterion scaling of questionnaire items. Socio-Econ. Plan. Sci., 1969, 2, 355-362.

Coleman, J. S., et. al. *Equality of Educational Opportunity.* Washington, U.S. Government Printing Office, 1966.

Green, R. L.; Hofman, L. J.; Hayes, M.E.; Morgan, R. F. The Educational Status of Children in a District Without Public Schools. Cooperative Research Project No. 2321, 1964.

Hawkridge, D. G.; Chalupsky, A. B.; Roberts, A. O. A Study of Selected Exemplary Programs for the Education of Disadvantaged Children. Parts I and II, September 1968 and Follow Up, June 1969. Palo Alto, Calif.: American Institutes for Research.

Horst, P. *Factor Analysis of Data Matrices.* New York: Holt, Rinehart and Winston, Inc. 1965.

Husen, T., et al. *International Study of Achievement in Mathematics.* Vols. I and II. New York: John Wiley and Sons, 1967.

Mayeske, G. W., et. al., *A Study of Our Nation's Schools.* Washington, U.S. Government Printing Office, 1969.

Newton, R. G. and Spurrell, D. J. A development of multiple regression for the analysis of routine data. Applied Statistics, 16 (1), 1967.

Plowden, B., et al. *Children and Their Primary Schools.* Vols. I and II. London: Her Majesty's Stationary Office, 1967.

Shaycoft, M. F. *The High School Years: Growth in Cognitive Skills.* Pittsburgh, Pa.: American Institutes for Research and School of Education, University of Pittsburgh, 1967.

Thorndike, R. L. *Personnel Selection.* New York: John Wiley and Sons, Inc., 1949.

Chapter 6

THE ASSOCIATION OF TEACHER RESOURCENESS WITH CHILDREN'S CHARACTERISTICS

Stephan Michelson

If we can arbitrarily, and without precise distinction, consider that schooling might affect skills, values, and personalities, there is a difference of opinion about which of these actually occurs: [NOTE: Throughout this chapter numbers in brackets refer to references at the end of the paper—Ed.]

> The school, then, is an organizational embodiment of a major social institution whose prime function is to bring about developmental changes in individuals. It is an agency of socialization whose task is to effect psychological changes that enable persons to make transitions among other institutions; that is, to develop capacities necessary for appropriate conduct in social settings that make different kinds of demands on them and pose different kinds of opportunities. [(9), p. 3.]

> As social scientists, we maintain a skeptical view concerning the efficacy of formal schooling for the teaching of values. To the social scientist a view of formal education as an omnipotent socializing agent shows an exaggerated regard for education. The social scientist is not convinced that institutions of formal education are capable of accomplishing all the mammoth tasks that some apparently expect of them. The classroom may well be a place where formal skills are learned; it may also contribute to the transition from the family to the larger society. Finally, it may contribute somewhat to the maintenance of a core

> culture or the creation of a cultural synthesis. But whether formal education really has much influence on either cultural values or social behavior is not evident. [(14), p. 7.]

The recent rapid entry of model-oriented social scientists, sociologists, and economists particularly, into educational research has brought an unfortunate emphasis on the latter point of view. Skills, being easily measurable, are taken to be the outcome of schooling in most statistical studies. An empirical approach not relying on statistical analysis led Dreeben to his conclusion. He observed the structure of schools, asked what that structure *could* produce. With Callahan's work [5] as additional evidence, one could conclude that the major outcome of schools has not historically been meant to be cognitive skills. And for purposes of generating income, the work of Gintis [12] and Berg [2] indicates that cognitive skills are not necessarily the most useful outcomes of schooling.

Nonetheless, recent investigations of school outcomes and the school characteristics that affect them (or do not affect them) have centered on these skills which schools may not have been intended to produce, are not structured to produce, and would not necessarily benefit people if they did produce. Studies continue, this one no exception, to ask questions about the relationship between inputs and outputs despite the fundamental lack of knowledge of what outputs are desired, possible, and efficacious.

The ideas set out here, the kind of research described, therefore must not be taken as evidence for one kind of school structure as against another. It is too facile—and too common—to investigate one area of school production, ignoring the consequences in other areas. It could certainly be that a technique, say tracking, did successfully increase cognitive skill acquisition at all levels, *and yet was entirely unacceptable as a method of school organization.* [1] Thus I will discuss the question of the specificity of teacher characteristics in producing outputs such as reading scores, or even student attitudes, without meaning to imply that if certain types of children respond better to different types of teachers, then the schools should be organized to match them. This will be one argument that some such organization might be desirable, but for many reasons it may not be. I will conclude the paper with a suggestion about a school authority structure which might better accommodate my findings and general theory. But this is meant to be tentative and suggestive, not persuasive. That is, there are two kinds of arguments against my findings: First, one could argue that they are incorrect or at least inconclusive. This is a technical kind of discussion which would hopefully result in the design of a test which would confirm or deny the results reported here. Second, one could accept my results, but reject their implications because the school policies they imply are unacceptable. I hope only to set the tone, and, I pray, a trend, that one cannot advocate school policy on the basis of a very limited set of school outcomes, say, on the basis of skill production, absent any knowledge of the personality or value system effects of that policy.

The Paper in Outline

With this brief caveat, I will here outline the intended progress of this paper. The next section begins with a limited discussion of school production, and discusses some characteristics which I deem important to an *ex post* cross section investigation of the effects of schooling. This discussion is intended to begin to clear the air about different conclusions which have been reached regarding the association of school and teacher characteristics with student test scores. The way to determine which study has reached correct statistical conclusions is to investigate the properties of the investigations: the samples, definitions of variables, statistical techniques employed. These must be justified, and the results of a study must be weighted by the appropriateness of the techniques.

Following this exposition, ordinary least-square estimates of the relationships between test scores and school inputs are presented and discussed. The interpretation of statistical results is a separate issue from their correctness, and my claims for my interpretation will be far more cautious than my claims for my findings. There, however, some basic points of this paper will begin to emerge. A brief exposition of a simultaneous equations system will add fuel to the fire.

In the third section, the implications which might be drawn from the statistical presentation are examined. Concepts such as "resourceness" and "specificity" will be defined in terms of the regression results. However, the inferences are tentative, and some ways in which they might be altered are suggested. I will conclude the paper, then, with a brief fourth section about the implications of this work and its tentative interpretation for school administration. A possible modification of the present structure is offered—as is the whole paper—as suggestive, not definitive.

I include, as an appendix, a review of some material from the field of teaching "exceptional children," especially the blind, deaf, and mentally retarded. The emphasis will be on the acceptance, in these cases, of the concept of teacher specialization by *type of* *child,* as opposed to specialization by subject matter. My major effort in the text of this paper is merely to extend that already accepted notion to a broader view of the need to consider the characteristics of the pupils in making teacher assignment, and in teacher training.

Statistical Investigation of Teacher Resourceness

The exposition here will not be abstract theory, but the theory which leads directly to use of the Equal Educational Opportunity Survey (1965) data to investigate the association of school and teacher characteristics with student outcomes. The exposition will discuss the following, in order: the data sample, the observations, the variables, and the statistical technique. Especially in the last section, I will assume that the reader is familiar with the paper, "A New Model of School Effectiveness" by Henry M. Levin [27] prepared for this book. The sample and variables used here are identical to those used by Levin, and the simultaneous model is similar.[2]

The Sample

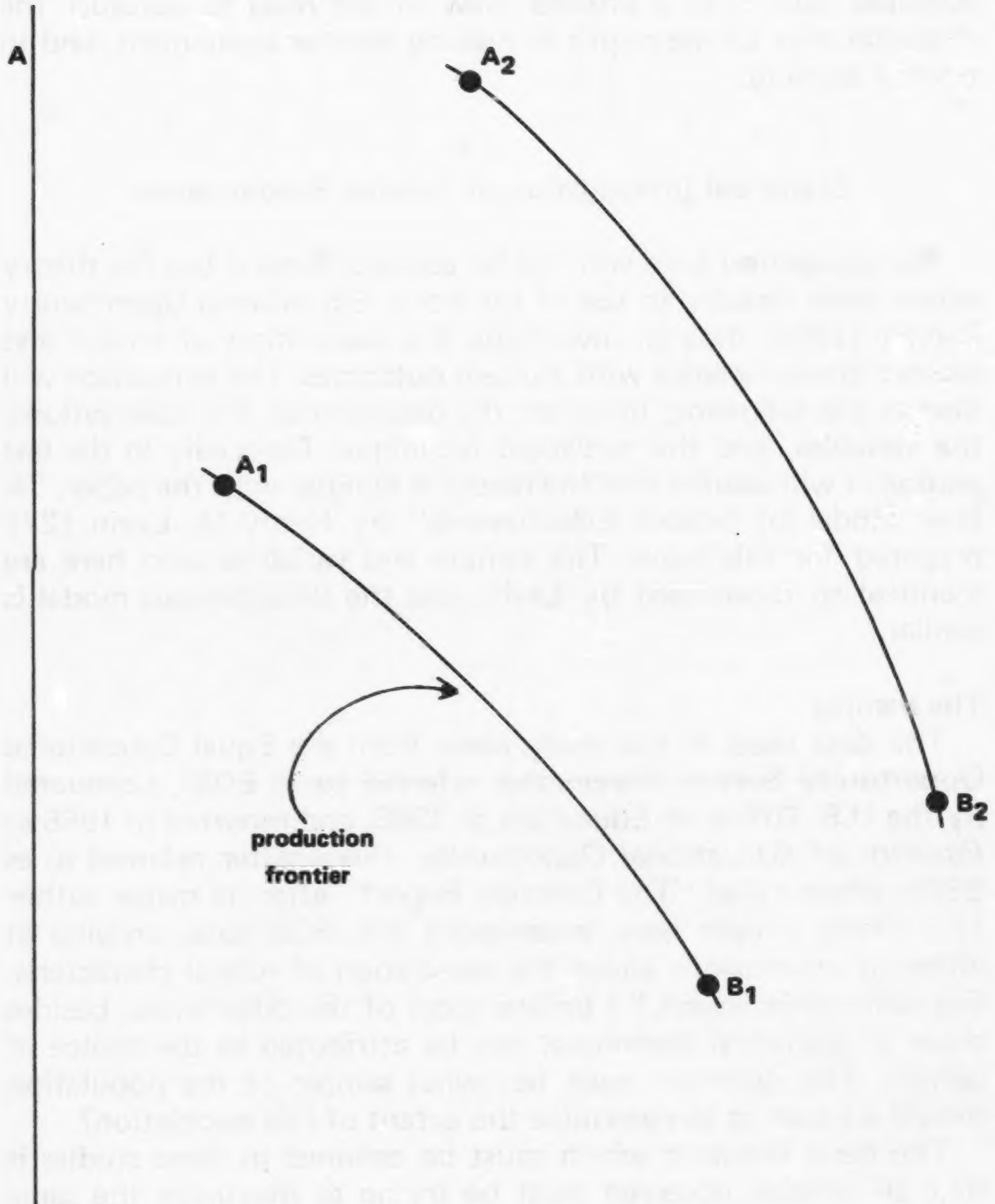
The data used in this study came from the Equal Educational Opportunity Survey (hereinafter referred to as EOS), conducted by the U.S. Office of Education in 1965, and reported in 1966 as *Equality of Educational Opportunity,* (hereinafter referred to as EEO), often called "The Coleman Report" after its major author [7]. Many people have investigated the EOS data, arriving at different conclusions about the association of school characteristics with achievement.[3] I believe most of the differences, besides those in statistical technique, can be attributed to the choice of sample. The question must be: what sample of the population should we look at to determine the extent of this association?

The basic constant which must be assumed in these studies is that all schools observed must be trying to maximize the same thing, hopefully our output measure, though that is not strictly necessary.[4] And they must be acting this way for *all children* in the school, or else we must observe only those children for whom this is true. Figure 1 shows the case in which two outputs, A and B, are related by the "production frontier" as indicated. This is merely the locus of possible outcomes with the resources at hand.[5] Schools A_1 and B_1 tend to produce A and B respectively, as do A_2 and B_2, which are endowed with more resources. The more resources of B_2 do not produce more of A than A_1, nor does A_2 produce as much B as B_1. We can find statistically that

388-306 O - 70 - 9

47

Figure 1



NOTE: A and B represent outcomes of schooling.

Each production frontier represents the locus of possible outcomes from the school resources. (2 indicating more resources than 1), depending on to what ends they are used.

resources do not affect either A or B, when in fact they affect both regardless of which is preferred by the school.[6]

Within a school district there is a variation of social class among schools which might lead to variation in aims of programs. There is also variation of class within schools which might induce differential program aims for different children. The same kind of variation in aim occurs among districts, but I think less of this variation occurs within than between districts. Many overt and covert policies of school boards which indicate differences in their aims can be controlled: the factory town which in general produces workers for the plant, the prestige suburb which produces college graduates, the central city which produces a spectrum and, like New York or Boston, allows its citizens to be chosen "fairly" (that is, by exam) into the prestige high schools. The aims of the school board, the environment of the city (air pollution, garbage collection, etc., all of which could have education consequences; even the mean temperature)—all of these variables are controlled by choosing one large city with several schools. This sample is not perfect: the dilemma of Figure 1 has not been solved.[7] But I believe it is considerably reduced. To the extent that this problem still occurs, the observed association between school characteristics and children's achievement is reduced below the actual association.

In addition, previous studies have included children in the sample who had not been in the same school in preceding years. They were identified with their correct home variables, but incorrect school variables. In many cases, this is probably not serious: some children transfer among very similar schools. By choosing a central city, the upwardly mobile children who have recently moved to the suburb are eliminated. Those who will soon move out may remain unidentified. However, the resulting bias in the association of school variables with output is toward zero, while not affecting the home variables. Although this bias is not unacceptable, it is not necessary. I have eliminated from the sample those children who had not been in the school in question since the first grade.[8]

The sample, then, comprises those children in a large eastern city, "Eastmet," who had attended only one elementary school. This sample was divided into whites, blacks, and others, only the white and black samples being utilized for this study.[9]

Observations

Debate among researchers has been endless about whether one ought to observe individual children or school means in this type of study. The question is often based on argument about the number of degrees of freedom when individual children are used: is the number of schools, or the number of children the base? I will surely not answer this question to the satisfaction of people who think differently, but explanation of my procedures follows.

Most of the variation in test scores occurs within schools. Children within schools differ more from each other than schools as groups do from each other. This is an interesting finding. It has been used to show that schools are relatively ineffective, for better schools should produce better students. However, since there is grouping within as well as between schools, there is no reason to believe that schools are ineffective on these grounds. We are back to the Figure 1 problem: if each school chooses some students on whom to stress the outputs we measure, others to stress other outputs, then schools could be totally effective, produce *all* variations, and yet there would be more variation within than between schools. Furthermore, if the selection were made by social class, then the social class variables would be associated with output differences.

To see this, consider several schools which are formed by random selection of students from a community. Within each school, children are grouped by their behavior, which is correlated with their social class. The more cooperative, passive students are put in the high "track," which stresses academic output. The lower tracks stress behavioral outputs more and more. By grade six, the upper track has been reduced in relative size by elimination of those who, though behaviorally adept, do not succeed academically. Lower track academic successes, however, do not move up.[10] The mean social class and mean test scores will be equal among schools. Within schools social class and test score will correlate highly. If one were bound to interpret "social class" as necessarily indicating *home* influence, and observed school means, he would conclude that schools had no effect. By construction, however, this conclusion would be incorrect.

In fact schools are not alike by social class or achievement. Some interschool variation is observed, and it correlates with social class more than with school characteristics. However, consider the other polar case: the tracking I described above now occurs *among* schools. School #1 is initially selected by social class, though by grade six some upper class children have been moved to schools #2 and lower. The interaction of high social class and reasonably high ability would perfectly predict placement in school #1, and therefore test score. By linear regression where only social class is entered, that variable would predict quite well. Since school resources in this case would be allocated by function—academic resources to the academic school, etc.—school variables would also predict outcome.

The facts seem to lie somewhere in the middle: schools are relatively homogeneous by social class, as in the first polar case, but not completely so. Since abilities vary within social class, and social behavior varies within each school, each school can have its academic, each its nonacademic group. The variation between schools, which would be greater if schools were treated as in the second polar case, is reduced by intraschool grouping. But some between school output variation still occurs, and it is associated with the mean social class of the school. The interpretation that it is therefore "due to" the social class of the school is correct, but the interpretation that this operates through home life of the students is incorrect. Similarly, when one finds that a lower class child does better academically in an upper class school, one need not conclude that this is due to the direct influence of his classmates on him. It may be that the school he is in stresses academic outputs more than schools with more of his social class equals. There is simply no reason to believe, from the correlation between social class and academic success, either by school mean or by individuals, that the cause of this association is the home life of the children.[11]

This argument, then, speaks to the issue of whether to observe school means or individuals in this sense: By the models just presented, the association is between the child and his output. To what extent this association is found between schools depends on the school structure, i.e., to what extent grouping occurs within or between schools. This extent may vary from city to city, and even within cities. It seems wise, then, to observe children directly.

There are other arguments: Children are of more interest than schools. I don't know what to make of the fact that mean school resources do not correlate with mean school output. The resources going to a child might still be very important. Since the variables labeled "school characteristics" do not vary within schools, obviously I cannot determine the effect of within school variation in these characteristics with these variables. But I can still pick up their effect to the extent that I can identify the individual characteristics by which these inputs are allocated. The problem is partly one of interpretation, and partly that the correlation between individual characteristics which we measure (which exclude, for example, direct behavioral measures) and the allocation of school inputs may not be perfect.

The variation which we want to explain, then, is variation in student scores, not variation in school scores. The fact that this variation occurs mostly within schools, that the percent of this variation which we can explain with the variables we have is small (about 47 percent of verbal score variation, 36 percent of reading score, for whites), is a fact not to be covered up by observing the relatively invariant school means.

The argument about degrees of freedom, in this context, is nonsense. We observe children in situations. There are not as many situations as children. But similarly there are, for example, only two sexes, nine categories of possessions, 50 possible scores on the verbal test. These numbers have nothing to do with degrees of freedom. When two children in the same school receive different test scores, then the association between the school characteristics and those scores is reduced. That is an accurate portrayal of the situation: knowledge of aggregate resources does not predict individual success. It is like observing the difference in behavior between married men and bachelors. If a thousand observations are taken, then the degrees of freedom calculation begins with 999 on taking the mean, and is reduced from that figure by adding independent variables. It is not *two*. To the extent that variations of behavior within the categories "married" and "bachelor" may occur, they indicate that this variable is not a good explainer of that variation. But the degrees of freedom are not affected by this consideration.

Suppose everyone who is married lives in a private home, and all bachelors live in apartments. Then entering type of living quarters would be redundant if marital status is already included. Similarly, if there are only 34 schools with whites in Eastmet, no more than 33 school variables can be entered into a regression equation. From the 34th on, each variable can be expressed as a linear combination of the others. But this does not limit the degrees of freedom when some small number of school variables are entered, any more than one would argue that there are only two degrees of freedom in an equation which contains only marital status, despite the fact that marital status and type of dwelling cannot appear in the same equation. In the white equations, 597 children are observed in situations in which the ordering of school variables is restricted. All children in school A receive all the inputs in school A, and those in B receive B. Not all possible interactions are directly observed in the data. This is typical of regression data—it is why regression analysis is used. The statistical degrees of freedom do not depend on the many possible (and redundant) variables which are *not* entered into the equation, but on the number of observations, less 1 plus the number of independent variables which are entered.

My argument, then, is that it is reasonable, preferable, and statistically valid to consider children as observations. It is reasonable and preferable because the object of the investigation is to determine the effects of variables on children, not on schools. It is valid because school variables act like any situation variables, and do not restrict the degrees of freedom of the equation.

The Variables

Data are from the sixth grade questionnaire, the teacher questionnaire, and the principal questionnaire, all of which are reprinted at the end of Volume I of EEO. I selected those teachers who were in the third through fifth grades, because the test was given in September of the sixth grade.[12] The teacher responses were averaged over the school, and the average was applied to each pupil in the school.

This procedure implies that each student moves randomly among teachers through the grades. For future researchers, a suggestion from Marshall Smith[13] is to weight each teacher by the percent white which he reports relative to the percent white in the school, and apply this weighted figure to white students, and apply the complementary weights to the teachers for black students. This seems to be a better approximation than mine to the data we all desire, but no one has: the correspondence of particular teachers with pupils through several grades. In either case, errors of association should bias significance tests, and possibly (if assignment is nonrandom) even the statistical relations between teacher characteristics and student outcome towards zero.

A recent study notes that "the evidence suggests that the quality of the principal and staff has a profound influence on [student] improvement." [(33), p. 1.] Though in EOS there was evidence on the principal's degree, major, and experience, there was no direct measure of the principal's performance (such as the 30 question test taken by teachers), or attitudes (such as teacher preferences for other school, for different race or "ability" of pupils). I therefore used only his answers to questions about the school, and not about himself.[14]

Individual student questions were sometimes combined, sometimes divided by possible answers, usually according to my judgment or interest, sometimes according to preliminary findings. For example, I started with a linear age variable, which associated negatively with output: the older the child, the lower the achievement score, controlling for other factors. But there was really no significant difference between a 10-year-old and an 11-year-old—and in fact, 9-year-olds (children who reported that they were 9) were below average. Thus I created binary coded variables for 12 or older, and 9 or younger. On the other hand, I combined nine home items into an index of possessions, not being ready to believe that the possessions of any one provided the information I was seeking.[15] The names attached to these variables should indicate how they were created.

For some of the equations to be presented, some interaction variables were also created. These were formed by visual inspection of school summary data. School resources and average student characteristics were looked at, where "resources" were average teacher test score and experience, and the pupil characteristics were possessions and a socioeconomic index.[16] At least four schools had to meet criteria of "low," "mid," or "high" socioeconomic status of the students ("peers"), or three categories of resources to qualify as a variable. Three categories of schools were selected this way: high resources but low peer, low resources but mid peer, high resources and mid peer. The effect of each of these categories was not assumed homogeneous, but was made into a separate variable for above median and below median SES for each child. The interaction effect of being a high SES child in a low SES school, or a low SES child in a low SES school could be accounted for separately. These interaction variables were not included in the simultaneous equation system.

The outputs considered are raw test scores of students. A verbal test was the basis of most findings previously reported. This test, and in addition a reading and a mathematics test, are used in the single equation study. In the simultaneous model, only verbal score is used as an academic output. An index of student attitude and his grade aspirations, are also outputs in the model. Grade aspiration means how far the student says he wants to go in school. However, 87 percent of the blacks in the final sample, and 93 percent of the whites had the highest two values among five possible values. The student attitude question on the other hand, was very evenly distributed. Of eleven possible values, between 10 percent and 20 percent of the blacks in the final sample had each of five values, and 10-20 percent of the whites had each of six values. It seems trivial to assume, but nonetheless important to mention, that high values of grade aspiration indicate "expected" or "socialized" response. The attitude questions, such as "If I could change, I would be someone different from myself" (answers "yes," "no," "not sure"), are not those ordinarily asked of a sixth grade pupil, and so elicit less socialized, more spontaneous responses.

Finally, I will touch here a little on interpretation of variables. The authors of EEO sagely warned about "the danger of unconsidered surrogates," which "can lead to seriously misleading conclusions." They give an example:

> Let us suppose that community attitudes toward the importance and quality of education have substantial effects on the development of student achievement. What would we expect about the apparent relation between achievement and teacher characteristics? Surely we would expect that communities more concerned with education and

> educational quality would—(1) be more selective in hiring teachers, and (2) pay higher salaries, thus attracting better candidates. As a consequence we might expect an apparent relationship between development of achievement and measurable teacher characteristics to be generated as a surrogate for an underlying relationship between development of achievement and community regard for education, even if teacher characteristics themselves had no effect on achievement. [All quotes, EEO, (7), p. 327.]

This warning is perfectly in order. The example, of course, does not apply in the present case, where one city only is being studied. Strangely, nowhere in EEO is the suggestion made that surrogates can work the other way round: that home items can be surrogates for access to school facilities. Take, for example, the problem of student assignment to teachers, mentioned above. Though there is some meaning to the average teacher characteristic in setting the atmosphere of the school, the deviation from that average which is each child's history may have a regular pattern. I have been told, for example, of a very aware teacher in a Boston suburb who takes her low-tracked class through the school corridors, looking into other classrooms. The students one by one mark, from visual observation through a window in the door only, which track each class is in.[17] Their estimates correlated well with the actual tracking, the identification coming, says my informant, from the *dress* of the children in each room. If teacher assignment among tracks is biased, and if the characteristic by which teachers are assigned to higher track students is truly effective, that effect will show as a student variable. It may be in the possessions index, size of family, father's education, mother has a job, etc., whatever correlates with type of dress.

In fact, in assessing the probable direction of surrogates, the side taken by EEO seems perverse. Only student characteristics vary within schools. We know that school facilities are not distributed randomly within schools, and any student variable which is associated with a bias in resource allocation may be a surrogate for the effect of that resource. There is no such striking argument on the other side, especially in a one city sample. One must assume that individual student items are more likely to be surrogates for school effects than vice versa.

There is no way to add the possible biases together to come with a resultant. However, I have attempted to bias all estimates away from finding that school resources are associated with the outputs. Other studies have been similarly biased, but they have either not recognized or not stressed this bias.

In interpreting the variables, the prime rule will be *a priori* to *suspect the label* of the variable. All schools probably track, so what the "tracking" variable indicates is something about the form of the tracking, the nature of the principal who decides which way to answer the question, a student body so homogeneous that tracking is not feasible, or something else. The teacher test, often taken by the teachers together, never under professional supervision, may indicate degree of cooperation among teachers. The number of library volumes is presumably an estimate from purchases or the card file, and not an indication of the actual number available for students, nor of course of their quality, the physical ease of taking them out, the extent to which students are introduced to the library, encouraged to use it, etc. Each item has the same interpretation problem.[18]

Statistical Techniques

The common technique applied to EOS and similar data is the single linear regression. A dependent variable is made a function of a set of independent variables, and fitted to the data to accord to the form:

$$Y = a + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n$$

The fit is made according to the principal of least squares, which minimizes the sum of the squares of the distance (in the Y direction) of the observations (data points) from the fitted n dimensional plane, where n is the number of independent variables. I assume that the reader is somewhat familiar with this technique. I will mention here that by minimizing the sum of *squares,* distant points receive a weight greater than the researcher would perhaps like to give them. They may be due to some different relationship—such as the desired output of the school, as discussed above—and should not be allowed to affect the estimates.

In using time series data, or other data with a limited number of observations, one often performs a residual analysis. War years, depression years (in time series), Alaska and Hawaii (in State observations), and other such identifiable anomalies from common patterns often cause the outlying points. Sometimes they are entered into the equation by creating special variables, sometimes they are excluded. In the case at hand, however, even if we did find one school or two with observations far from the rest, we would not know why this was so. If we did know, it would be because we had a variable describing those schools which had different values for them, in which case inclusion of these variables should solve the problem. Hawaii and Alaska are often different from the other States because the meaning of "nonwhite" in, say, generating income, is different in these States from those in the 48 other States. Rationing in war years made the notion of "price" different from ordinary years, and the composition of output, demand for labor, etc., were unusual. A dummy variable in these cases corrects, from information external to the data, for a variable in the data which has different meanings over different observations.

Not knowing which Eastmet schools are which, not having any information about them individually outside of the data, a dummy variable for certain schools would only be a measure of ignorance in an effort to improve R^2 or other measures of goodness of fit. It might be an interesting investigating device, but not an explanatory device. On the other hand, as explained above, I did pick out some combinations which could lead to extreme observations, and defined variables accordingly as "interaction" variables. Their purpose is to bring extreme points into the general scatter, to reduce their influence on the resulting coefficients. The coefficients of these variables themselves are not interesting in this context.

There are a number of basic problems with the single linear regression. One is in its use: It does not, and cannot in simple application be a description of the *production process* within schools. A process should be described before being estimated, and I cannot believe that anyone would describe the schooling process as linear additive. Surely there are many interactions, many nonlinear effects. One might be able to estimate them by linear regression on a reduced form model, deduced from a series of equations describing student preparedness, teacher ability, desire, etc. I have not seen such an attempt made.

What a linear regression on the variables might do is give coefficients which describe in some average way the effect of the independent variables on the dependent variables. The production function must be correct *on the margin*: it should predict what an increment of X_i will do to Y, holding the other X's constant.

The linear equations presented here and elsewhere in the education literature should not pretend to do this. They perform, rather, an averaging function. They designate what the linearly isolated effect of a particular variable seems to be; at least, what the linearly isolated association of an independent variable with the dependent variable is over a large number of observations. If there is a large coefficient for an inexpensive variable, the linear regression does *not* imply that more of that variable should be purchased. On the margin, that variable may have little effect.

A regression estimate fits the scatter of observations such that it is the variations in the observations which create the hyperplane, not their levels. One problem in interpreting the results of average equations is in determining the effect of variation in inputs relative to their base. Explanations of variations in scores are not explanations in levels. Most students in our total sample scored 30 or better out of 50, and *all* students scored 20 or better in the verbal test. Most of the questions had five possible answers, so pure guessing would have produced a mean of at most 10 correct answers.[19] The worst student did twice as well as that, and the average student did three times as well. This does not indicate that schools, as opposed to home life, produced this level of achievement, but it is possible that, at least for some children, schools did perform this function. The variation in school resources may produce little of the variation in outcome, but the existence of schools might produce most of the test score level—or none. That is still an open question.

The single equation linear variable cannot account for the effect of attitudes on achievement, if attitudes are also the result of achievement. Simultaneous determination of attitudes and achievement requires a simultaneous equations model. The three equation model presented here is a variant of that employed by Henry M. Levin [27] in his paper for this conference, and I will not go into detail about it here. Student's grade aspiration and "fate control" attitude are assumed functions of the same variables as his achievement, *and also a function of the achievement itself.* Achievement is also a function of these attitudes. Three equations containing arguments which are dependent variables elsewhere in the system, must be estimated by two-stage least squares. The model is overdetermined *a priori.*

The Equations and Their Implications

In this part of the paper I will present regression equations derived from the Eastmet samples. In the first section, the ordinary least squares "average effect" equations will be presented and briefly discussed. Hazards of interpretation will be stressed. In the next section the equations for blacks and whites will be compared with each other to see if the same equations describe the average effect of the variables on different children. In the third section, equations for whites will be compared by social class. Finally, a simultaneous equations system is presented and compared by race.

Average Effect Equations

The average effect equations, as explained above, are regression estimates of the average relationship between the dependent variables (verbal, reading, and math scores)—one at a time—and student background, school and teacher variables, with some attempt to account for points far from the resulting hyperplane. They are not attempts to describe the production process where the independent variables are "inputs," the dependent variables "outputs."[20] I do not feel constrained to choose a "best" equation for each output, but will present alternatives when no clear choice can be made.

With this kind of data, the crude measurements, the many possible interpretations for any variable, this freedom is advantageous. For the white sample, two equations with verbal score as the dependent variable, three with reading score, and two with math score are presented in table 1. For blacks, two verbal, two reading and one math equation are presented in table 2. The fewer black equations is a manifestation of the common finding that black behavior and outcomes are not as associated with typically measured variables as white behavior and outcomes. This is because we measure the wrong variables for blacks, their behavior is erratic with respect to the variables, and society's behavior is erratic with respect to the variables when dealing with blacks. By measures of goodness of fit also, the black equations do not explain as much of the variation in scores as do the white equations.[21]

The different specifications of equations generally contain the same student variables, substitutions being made among teacher and school variables. Sex and age were included *a priori,* and possessions and size of family, the most significant variables in almost every equation,[22] were included essentially automatically. The other variables were experimented with, but the bias in selection was to include as many student variables as possible. There is therefore a bias against the inclusion of school and teacher variables, so that there is no question about their appropriateness in these equations.

An example of the distinction between the average effect equations as presented here and production estimations can be drawn from the "kindergarten" variable, which appears positively wherever it is included. This does not indicate that sending a child to kindergarten will raise his sixth grade verbal score by over two points (if he is white). It indicates that white children who went to kindergarten scored, on the average, two points higher on this test than other white children with otherwise similar characteristics. The kindergarten may or may not have played a role in this higher score; it may indicate the concern of his parents, or the neighborhood in which the family lived, or their social milieu (in which it was understood that children went to kindergarten before

TABLE 1
AVERAGE EFFECT EQUATIONS, WHITES

Independent Variable	Verbal$_1$	Verbal$_2$	Reading$_1$	Reading$_2$	Reading$_3$	Math$_1$	Math$_2$
Constant	17.5 (5.4)	−16.9 (3.1)	.6 (.2)	-5.8 (.8)	10.1 (1.9)	5.8 (3.8)	10.5 (2.9)
Background:							
Sex	.8 (1.3)	.8 (1.4)	1.8 (3.9)	1.8 (3.7)	1.8 (3.8)	−.7 (2.0)	−.8 (2.3)
Age 12+	−7.5 (5.0)	−7.4 (4.9)	−4.6 (4.0)	−4.5 (3.9)	−4.7 (4.1)	−3.0 (3.4)	−3.0 (3.5)
People at Home	−.5 (3.3)	−.5 (3.4)	−.5 (3.8)	−.4 (3.7)	−.5 (3.9)	−.3 (2.3)	−.2 (2.3)
Possessions	1.4 (5.3)	1.3 (4.8)	1.0 (5.2)	1.0 (5.1)	1.0 (5.0)	.3 (2.2)	.3 (2.0)
Father's Education	.3 (2.7)	.3 (2.9)	.3 (2.9)	.3 (2.9)	.2 (2.6)	.3 (4.6)	.3 (4.4)
Kindergarten	2.1 (2.3)	2.2 (2.5)				1.3 (2.5)	1.2 (2.3)
Mother's I.D.						-.3 (1.5)	
Teacher:							
Test Score		.8 (3.2)		.5 (3.0)			
Experience	.4 (5.1)	.6 (8.2)	.4 (6.5)	.4 (6.4)	.1 (1.4)		
Tenure					−2.0 (1.2)		−2.6 (2.5)
Race Discrepancy	−2.7 (5.2)						−.3 (1.1)
Race Preference		1.5 (4.5)	1.1 (4.6)				
College Major							2.6 (2.0)
School Tracking				−.6 (1.9)	−.5 (1.4)	−.7 (2.7)	−.5 (1.9)
Library [1,000's]	.3 (2.1)					−.6 (2.5)	
Aud.-Caf.-Gym	.7 (3.2)			.3 (2.2)	.3 (1.9)	.3 (2.8)	.2 (2.0)
Acres						.6 (3.2)	

(Table 1 -- cont.)

	$Verbal_1$	$Verbal_2$	$Reading_1$	$Reading_2$	$Reading_3$	$Math_1$	$Math_2$
% Upper Quars. x 10^2					6.6 (3 .2)	5.9 (4.7)	5.9 (3.4)
Interactions:							
HiWh.–LoNW	–3.6 (2.2)					1.8 (2.2)	
HiSES–LORes–MidPr	2.7 (1.5)	5.8 (2.9)	2.3 (1.7)	4.5 (2.9)	1.8 (1.2)		
LoSES–LORes–MidPr	–3.6 (1.7)		–2.8 (1.7)		–3.5 (2.0)		
LoSES–HiRes–LoPr	–7.4 (2.5)	–6.1 (2.1)					
R^2 [corrected]	.476	.470	.361	.353	.358	.333	.327
S.E.	7.305	7.347	5.666	5.699	5.678	4.181	4.200
Deter.	.140	.222	.492	.246	.033	.0178	.0696

NOTE: T statistic below coefficient refers to coefficient = 0.

TABLE 2
AVERAGE EFFECT EQUATIONS, BLACKS

Independent variable	$Verbal_1$	$Verbal_2$	$Reading_1$	$Reading_2$	$Math_1$
Constant	1.4 (.2)	2.7 (..3)	2.8 (.7)	5.1 (1.2)	7.2 (1.4)
Background:					
Sex	.7 (.9)	.8 (1.0)	1.2 (2.0)	1.2 (2.1)	-.6 (1.5)
Age 12+	–3.7 (2.6)	–4.0 (2.8)	–2.8 (2.8)	–2.8 (2.8)	–2.0 (3.0)
Age 9–	–4.2 (1.2)	–4.5 (1.3)			
Possessions	1.0 (4.0)	.9 (3.9)	.2 (1.3)	.3 (1.5)	.2 (1.8)
People at home	–.4 (2.4)	–.4 (2.5)	–.3 (2.4)	–.3 (2.3)	
Father's Education			.3 (2.8)		.2 (2.5)
Father's Occupation			.4 (1.6)	.4 (1.7)	.3 (2.1)

(Table 2 – cont.)

	Verbal$_1$	Verbal$_2$	Reading$_1$	Reading$_2$	Math$_1$
Mother's Education	.5 (2.9)	.5 (2.9)			
Mother's I.D.			–.5 (2.4)	–.5 (2.3)	
Kindergarten			1.2 (2.0)	1.2 (2.0)	
Teacher:					
Test Score			.2 (1.1)		
Race				–2.0 (1.5)	
Parents' Education	.8 (3.1)	.5 (1.9)	.3 (1.8)	.4 (2.8)	.2 (1.9)
*Years of School					–1.5 (1.5)
*Academic Major			–7.1 (2.1)	–6.3 (2.1)	
Tenure	–1.1 (2.2)				
School:					
Adequate texts	2.5 (2.0)	2.1 (1.7)			
Tracking	–3.3 (1.9)	–1.6 (2.8)			
Building age		–.05 (1.8)			
Library [1000's]		.6 (1.6)			
*Assignment					.8 (2.1)
Interactions:					
*Hi SES Hi Res–Mid Pri			7.6 (2.4)	6.7 (2.1)	
R^2 [corrected]	.152	.155	.132	.134	.074
S. E. of estimates	8.77	8.76	6.10	6.10	4.12
Determinant	.638	.343	.436	.638	.780

*Indicates no prior hypothesis about the sign of the coefficient.
T statistic below coefficient refers to coefficient = 0.

elementary school). If the children who went to kindergarten are different from those who did not, then no claim is made that the effect of sending a different child to kindergarten would be to add two points to his score.

The same distinction must be made for the teacher and school variables. For example, in the white verbal equations, the average discrepancy (per school) between the teacher's reported percentage of white students and *desired* percentage of white students is strongly associated with the score of the children if the teachers' average test scores are not in the equation. When we account for the test score, then not the discrepancy, but the absolute preference for whites has a strong effect. Verbal$_2$ surely does not mean that we should take teachers with mean test scores and consider those with strong preferences for white students to be the better teachers. If we did, we might then send them to schools where there are many blacks, where their discrepancy is high, and where they are consequently bad teachers.[23] Or we might find that these characteristics alone make no difference at all, on the margin.

What these coefficients probably mean is one of two things: (1) teachers are found to move towards their preferences, and white children who score higher tend to move toward whiter schools, so that teachers with strong preferences for whites tend to reduce their racial discrepancy and be associated with better students; (2) some teacher attitude, which may find some expression in racial preference, affects their teaching.

No policy conclusion follows from either interpretation, though the latter indicates that an area of investigation might be revealing: the effect of teacher attitudes on student performance. Some work on this question is being done, as is well known.[24] Whether the attitudes involved are trainable or selectable, whether they can be applied to all children in a classroom or by definition select within a classroom; to these questions I have no answers. And of course, whether these equations imply an effect of these attitudes on children or on teacher location is also open to investigation.

Comparing Equations by Race

It is not clear why, if the school variables are to be interpreted as social class phenomena, the black equations look so different from the white equations. The teachers' parents' education is an important variable in the black equations, but does not enter the white equations. Academic majors (as opposed to education or physical education majors) are negatively associated with black reading scores, but positively associated with white math scores. Teacher experience does not help black children—at least not experience in the teachers blacks have—and the race variable in the black Reading_2 equation substitutes for the test score in the Reading_1 equation, whereas neither variable appears in two of the three white reading equations. This is a serious question, to which there are several possible answers.

Blacks, it might be argued, are not able to gain resources by improving their social class status. [See Michelson (31)]. The phenomenon of the teacher associating himself with better students does not occur among blacks, possibly because housing discrimination is so strong that upper class blacks do not have access to upper class schools. Thus the association of quality teachers with quality students, which is the explanation behind the equations—this argument continues—does not apply to blacks, and the school and teacher variables which appear in the white equations have no chance of appearing in the black equations.

This argument is more incorrect than correct, though it probably has some of both elements. In my recent publication cited above, I presented resource indexes derived from some of the equations of tables 1 and 2. "Resources" were defined as those school and teacher items which appeared in the equations. Black resources were therefore different from white resources, and black resources were not distributed to blacks over social class, but white resources were so distributed among whites. However, whites' resources are also distributed by social class *among blacks.* There *is* an association between the average characteristics of schools and social class, when these characteristics are the variables entered in the white equations, whether white or black students are considered. These variables *could* have been associated with scores of blacks, which are also associated with social class (though not as strongly as white scores). But they were not. Instead, different variables appeared to be associated with black scores, and these variables were *not* distributed among blacks (or whites!) according to social class.[25]

A different argument, which accords with the allocation of these items, is that different things affect blacks and whites. That is, a characteristic of a teacher may be a resource for a white child—i.e., would increment his score—but not a resource for a black child. "Resource" then is not just anything which appears in a school, but an input which has an *effect.* What is a resource to whom is an empirical question. That question is not answered here, as I hope I have made clear. But it is raised here. It implies that the equations indicate some sort of causal relationship between *something* measured by some of the variables, and academic achievement. We do not know what that something is, because the variables are simply not that precise. But if there is any implication of causality in these equations, the implication should be stretched to include differential causality: different things affecting white and black children.[26]

At this point I have indicated that blacks and whites seem to respond to school variables differentially—i.e., that different variables have different resourceness to blacks than to whites. To indicate that this difference is statistically significant, I estimated the coefficients which blacks have for the white specifications, and the coefficients for whites with the black specifications. I then tested to see if their responses were the same. This is equivalent to asking if, with respect to these equations, blacks and whites could be said to be drawn from the same population.

In table 3, the F test, degrees of freedom, and significance level are given for all of the average effect equations. The conception behind this statistical measure is simple. The regression equation is estimated to minimize the sum of squared residuals: if Y is the test score and $\hat{Y}$ is the equation's estimate of the test score, then define $e = Y - \hat{Y}$. Minimizing Σe^2 is the same as minimizing $\Sigma \frac{e^2}{k}$ where k is any constant. If k is the number of observations (actually the number of degrees of freedom), then this expression is essentially the average value of a residual. If the average squared residual value is lower for separate regressions on subsamples than for the sample as a whole—it can never be higher—then the equations which generated these average squared residuals must be different. This will almost always be true to some extent, but since the average residuals from samples from the same population follow a known probability distribution (Chi^2), so does their ratio (F), and we can calculate if the reduction in average residual squared is statistically significant, i.e., highly improbable under the assumption that the samples were truly from the same population.

There should be no question that the blacks and the whites form two distinctly different samples. In fact, since most of the variables are the same in black and white equations—the background variables—this is a somewhat weak test. Further investigation of the individual school coefficients verified that they are different for blacks and whites under similar equation specifications. The *educational* impact of the *statistical* difference in resourceness cannot be so easily tested. This will be discussed below. But the point should be clear: the school variables which seem to be resources are different for blacks and whites.

Social Class Differences in Resourceness

Whites were split into bottom quartile and the rest, and the same test was performed. The results appear in table 4. Here, however, a few more words on the regression sample should be offered. In quartiling the sample by social class, the entire SMSA

TABLE 3

AVERAGE EFFECT EQUATIONS
F Test of Black–White Differences

	F	d.f.	Sig. %
WHITE EQUATIONS			
Verbal$_1$	4.07	14,1027	1%
Verbal$_2$	5.54	11,1033	1%
Reading$_1$	6.00	9,1037	1%
Reading$_2$	3.68	11,1033	1%
Reading$_3$	2.59	12,1031	5%
Math$_1$	3.99	13,1029	1%
Math$_2$	2.71	12,1031	5%
BLACK EQUATIONS			
Verbal$_1$	9.81	10,1035	1%
Verbal$_2$	9.69	11,1033	1%
Reading$_1$	4.73	12,1031	1%
Reading$_2$	4.14	12,1031	1%
Math$_1$	9.18	8,1039	1%

TABLE 4

AVERAGE EFFECT EQUATIONS
F Test of Bottom vs. Top Three Quartiles–Whites

	F	D.F.	Significance Level
Verbal$_1$	.97	14,569	n.s.
Verbal$_2$	1.20	11,575	n.s.
Reading$_1$	2.49	9,579	10%
Reading$_2$	2.16	11,575	10%
Reading$_3$	2.34	12,573	5%
Math$_1$	.048	13,571	n.s.
Math$_2$	.73	12,573	n.s.

sample was included. Though I doubt the representativeness of the suburb sample, together with the city sample I had a much more representative picture of class variation. In selecting the central city to study, a bias towards lower classes was produced. That is, more than one fourth of the city sample is in the bottom quartile. However, in selecting the sample of children who had been in one school only, the opposite bias was produced. I had no *a priori* expectations as to the result, but in fact only 32 of the 597 whites in the regression sample (5.4 percent) were in the bottom quartile sample. They therefore could not represent the entire spectrum of schools, though bottom quartile children are probably not in every school anyway.

In interpreting the results of table 4, the sample problem must be kept in mind. Difference in equations could be due to nonlinearities in the relationships, not differences in the sample, if the 32 children here represent extreme observations.

The Reading equations are apparently different. The coefficients were strikingly different for the bottom quartile regressions, including reversed signs for racial discrepancy and preference variables in all four equations in which these variables appear.

I partitioned the white sample again at the midpoint of the second to bottom quartile, creating a new lower sample with 88 (14.7 percent) observations. This adds more children to the bottom sample than were originally in it—and also undoubtedly adds more schools. Three of the four above-mentioned reversed signs reverted back to the signs from the total sample regressions. The R^2, which had been extremely high in the bottom quartile sample (above .7) went down (though were still high compared to the total sample R^2), and not one F test for difference proved significant. Once again, this could be a function of the particular schools involved. But it could also indicate that the bottom 5 percent of the regression sample children are very different in their reactions to school (and background) variables from the rest of the population, whereas the bottom 15 percent are not. Whether this means the bottom quartile of the entire sample is different, I do not know, and cannot determine from this data. None of these results can do more than suggest what may be true. But I think this kind of result is striking in educational possibilities, if not in statistical definitiveness.

The Simultaneous Equations System

The schooling process is not as simple as a single linear regression would indicate. One way in which to conceive of it is as a system which simultaneously determines several outputs which affect each other. As long as each output has determinants which are unique to it, such a system can be estimated. I propose a three equation model in which verbal score, student attitude (control over his life), and his grade aspiration are three outputs.[27] His attitude and his grade aspiration are functions of his score, in that they give him a sense of reality about himself.[28] Neither his attitude nor his grade aspiration influences the other directly, though they both influence the verbal score, hence each other indirectly.

Most of the background variables are assumed to influence all three outcomes, though whether the parents are "real at home" or something else (say, an uncle or aunt for father or mother) is assumed to have no direct effect on verbal score. Of the school variables, the teacher attitude question (preference for another school) is assumed to affect only attitudes and grade aspirations. Attitudes are affected by teacher turnover (principal's response to the question "What percentage of your teachers quit last year?") in that teachers in a school with high turnover might not pay as much attention to an individual as teachers in a low turnover school. Disruptions from turnover, and the other teacher and school characteristics (except teacher preference) all affect verbal score directly. The teacher's undergraduate institution was assumed to influence grade aspiration, though in this case (and this case only) the sign of the coefficient in the equation for whites was other than expected.

This three equation system looks like this:

$$V = b_1A + c_1G + \sum_i d_{1i}X_i$$

$$A = a_2V + \sum_i d_{2i}X_i$$

$$G = a_3V + \sum_i d_{3i}X_i$$

where V is verbal score, A is attitude, G is grade aspiration, X are the exogenous variables, and there is at least one $d_{1k} = 0$, $d_{2h} = 0$, and $d_{3j} = 0$, where $k \neq h \neq j$. In vector form, where Y is the output vector and X is the vector of exogenous variables,

$$Y = MY + NX$$

In this system, M is a 3 x 3 matrix, N is a 3 x 17 matrix, and Y and X are vectors with three and 17 cells. The solution is:

$$Y = (I - M)^{-1}NX$$

The structural equations are estimated by 2-stage least squares, and are given in tables 5 and 6 for whites and blacks, respectively. The solution, or reduced form equations, is given in tables 7 and 8.

TABLE 5

STRUCTURAL EQUATIONS, WHITES
N = 597

		TSLS	
	Verbal	Student Attitude	Grade Aspiration
Verbal	——	.054 (1.97)	.067 (3.34)
Student's Attitude	2.391 (1.62)	——	——
Grade Aspiration	1.622 (1.63)	——	——
BACKGROUND			
Sex	–.467 (.42)	.550 (3.08)	–.125 (.94)
Age–12^{+}	–5.026 (2.61)	.122 (.25)	–.284 (.79)
Family Size	–.080 (.29)	–.129 (2.49	–.048 (1.27)
Possessions	.630 (1.41)	.151 (1.57)	.021 (.29)
Kindergarten	.969 (.77)	–.116 (.41)	.579 (2.78)
Mother ID	——	–.021 (.18)	–.219 (2.45)
Father ID	——	–.091 (1.34)	–.051 (1.01)
Father's Education	.066 (.33)	.084 (2.13)	.017 (.59)
Mother has job	——	–.293 (1.45)	.305 (2.04)
SCHOOL			
Teacher Test Score	.246 (.96)	——	——
Teacher's Undergraduate Institution	6.457 (2.27)	——	–.349 (.80)
Teacher's Experience	.637 (5.10)	——	——
Teacher's Preference for Another School	——	–.147 (.37)	.701 (2.42)
Teacher Turnover	–.023 (.19)	–.048 (2.74)	——
Volumes Per Student	.380 (1.08)	——	——
Constant	–33.55	5.514	8.774
R^2	.364	.184	.254
S.E. of Estimate	8.144	2.163	1.603

TABLE 6

STRUCTURAL EQUATIONS, BLACKS, WHITE SPECIFICATION

N = 458

		TSLS	
	Verbal	**Student Attitude**	**Grade Aspiration**
Verbal		.072	.059
Student's Attitude	3.33		
Grade Aspiration	.048†		
BACKGROUND			
Sex	–.481	.199†	.551*
Age–12$^+$	–2.160†	–.210*	–.421
Family Size	–.395†	.032*	.019*
Possessions	.947	–.022*	.067†
Kindergarten	.253†	.017*	.793
Mother ID	——	–.089†	–.034†
Father ID	——	.050*	.085*
Father's Education	–.084*	.097	.098†
Mother has job	——	.001*	–.077*
SCHOOL			
Teacher Test Score	.254	——	——
Teacher's Undergraduate Institution	–1.463*	——	.675*
Teacher's Experience	–.179*	——	——
Teacher's Preference for another school	——	–.136	.960
Teacher Turnover	–.016	–.025	——
Volumes per student	.076†	——	——
Constant	–8.578	5.326	5.833
R^2	.146	.082	.194
S.E. of Estimate	10.36	2.179	1.992

*Black and white coefficients differ in signs

†Value of black coefficient more than twice or less than one half of the white coefficient.

TABLE 7

REDUCED FORM EQUATIONS, WHITES

	REDUCED FORM		
	Verbal	Student Attitude	Grade Aspiration
BACKGROUND			
Sex	.846	.595	−.068
Age−12⁺	−6.806	−.243	−.739
Family Size	−.613	−.162	−.089
Possessions	1.344	.223	.110
Kindergarten	2.135	−.002	.721
Mother ID	−.532	−.050	−.254
Father ID	−.395	−.112	−.078
Father's Education	.385	.104	−.043
Mother has job	−.270	−.308	.287
SCHOOL			
Teacher Test Score	.323	.017	.022
Teacher's Undergraduate Institution	7.718	.414	.167
Teacher's Experience	.835	.045	.056
Teacher's Preference for Another School	1.030	−.092	.770
Teacher Turnover	−.181	−.058	−.012
Volumes per student	.498	.027	.033
Constant	−8.030	5.084	8.237

TABLE 8

REDUCED FORM EQUATIONS, BLACKS

	Verbal	Student Attitude	Grade Aspiration
BACKGROUND			
Sex	.277†	.219†	.568*
Age–12^{+}	–3.808	–.485†	–.647
Family Size	–.382	.004*	–.003†
Possessions	1.159	.062†	.136
Kindergarten	.461†	.050*	.820
Mother ID	–.395	–.118†	–.057†
Father ID	.227*	.067*	.099*
Father's Education	.322	.120	.117†
Mother has job	–.002†	.0005*	–.077*
SCHOOL			
Teacher Test Score	.336	.024	.020
Teacher's Undergraduate Institution	–1.891*	–.136*	.563†
Teacher's Experience	–.237*	–.017*	–.014*
Teacher's Preference for Another School	–.540*	–.175	.928
Teacher Turnover	–.133	–.035	–.008
Volumes Per Student	.101†	.007†	.006†
Constant	12.497	6.228	6.573

*Black and white coefficients differ in sign.

†Value of black coefficient more than twice or less than one half of white coefficient.

I have not performed any statistical tests on these equations. Nonetheless, looking at the differences by race, the impression is strong that these are not the same systems. The number of different signs is striking. The specification was partly *a priori,* partly experimental. It was, however, perfected on the white sample.[29] Thus I could have derived an optimal black system, and asked what the coefficients for whites were like in that system, analogous to the work in the previous section. For the purposes of this exposition, the work presented here should suffice.

Interpretation of Statistics and Beyond

Some school inputs might be resources to some children, not to others. But this "all or nothing" approach to resources probably does not describe most of the things which affect children. Nor, of course, does it adequately account for the output problem: that what is an important resource for one output may be less of a resource for another, and may even have a negative effect on some objectives of schooling.[30] It seems easy to me to use the word "resourceness" to indicate that children respond to an input, realizing that some inputs have more resourceness (for some outputs) than others. Those inputs which have no resourceness are not resources, just as materials vary in their fluidity and those which have none are not fluids.

There are a number of ways to determine how much is "a lot" in terms of resourceness. Those items which have no statistically significant resourceness were generally excluded from the equations.[31] Besides statistical significance, one should consider the concept of educational significance. For example, the teacher test score for the one black equation in which it appears, Reading_1, has a coefficient of .2. We could ask: how many points would a teacher have to gain on his test score to raise the reading score one point, or one standard deviation.[32] Obviously 5 teacher points are required, on the average, to produce a point of reading score. The mean teacher test score for blacks is 22 points, and the highest possible is 30 points. Thus, as far as we can discriminate by this test, the best teacher would produce, on the average, 1.6 points more than the current average teacher. The difference between the average black and the average white reading score for the sample is 5.7 points.[33] Thus the experiment of putting the "best" teachers with the blacks reduces the black-white gap by 28 percent. On the other hand, calculating the black score if they had teachers with average test score equivalent to that of teachers of white children, 8.8 percent of the student score gap is closed. Both of these seem to be educationally significant.

On the other side, one might care more that these increases are 24 percent and 7.5 percent of a standard deviation, respectively, which might seem less significant. Another way to look at it is by asking how many whites score above the black mean, and how many whites would the black mean surpass under various assumptions. If the scores are normally distributed, then in the case where the means were equal, 50 percent of the whites would score above the black (= white) mean. Taking the white standard deviation and maintaining the normality assumption, then 78.5 percent of the whites score above the average black. Under the most favorable assumption, teachers who score 30 points assigned to blacks, but white teachers staying as they are, then 71.4 percent of the whites would still be above the black mean. With equal teachers, 76.3 percent of the whites would still be above the black mean. That is, for each 1,000 whites, 785 now score above the black reading mean (as opposed to 500 if blacks and whites were equal), and with "equal" teachers, the black mean would surpass only 22 more whites; with the best teachers, the average black would surpass 71 more whites (or 49 more than with equal teachers). One might consider these numbers educationally insignificant.

I see no unique measure of educational significance. Much of the question about the effect of variables is, like many other educational questions, a social problem, not a scientific one. Do blacks care more about their mean score relative to whites, or the number of whites who score better? I do not pretend to know.

Implications for Teacher Training

To this point, no inferences have been drawn from the statistical study to questions of policy. Two major areas of concern here are: teacher training and resource allocation. For this conference, the stress will be on teacher training.

The equations do not indicate that "resourceness" is a trainable phenomenon. Nor, assuming that to some extent it is, are the implications for training clear in terms of the content of any program. I have often thought that the Peace Corps and VISTA were excellent training for teaching, and several school districts have begun to think the same thing in the past few years. It does not seem to me to be necessarily true that school is a good place to train teachers.

Whatever the outputs desired, whatever the ways to train teachers to induce these outputs in children, what the foregoing does imply is that the structure of the training must respond to differences in the children who will be under the teacher's care. The concept that teacher resourceness differs by type of child I call "teacher specificity." Since different students will respond differently to different styles, attitudes, activities, language, strictness, etc., these properties of teacher activities should be investigated and directed to teachers who need them.

The concept of teacher resource being a function of the children being taught might lead one to conclude that segregated teaching was a preferred school structure. If this were so, one could still reject it, as I indicated at the beginning. But it leads to no such place. There are two obvious reasons why teacher specificity does not imply segregation.

First, other children may well be resources in addition to teachers. Teacher resourceness is not the only item in the entire resource package. Again, we don't know to what extent other children influence any particular child—nor do we know *which* other children influence any one.[34] But in this ignorance, to structure the schools by teacher resourceness would be to assume that other children have no effect. Even if this were true, the *fact* of separation (and the inevitable invidious comparison) is believed to have a detrimental effect on some of the children. Thus ignorance of the resource effect of children on children should, if anything, lead to more heterogeneous classes.

Secondly, teacher specialization itself need not lead to separation of children because that specialization may be different for different outputs. By and large, some teachers are probably better with underprivileged children, others better with overprivileged children. To that extent, they may go to schools which are also characterized as under or overprivileged. But some combination of resources may work best in a heterogeneous setting. That is, the specialization of some resources might be directed more at "mixed" children, whereas other resources might better be directed at one group or the other.

All of this is a land of mystery. Some teachers' talents are clearly in bringing diverse groups together, and other teachers are incapable of that. Some teach better with strict discipline, others with more freedom. Some have a conceptual approach to mathematics, some a mechanical approach. Some teachers will interpret Hamlet as weak, some will stress that he was tormented. Some are verbally oriented, communicate by words. Others prefer to play physical games, construct things. Some want to direct the class according to plan, some want to develop the sense of planning and conclusion seeking in children. Too much the search has been to differentiate between these characteristics in a search for the "right" ones. It seems strikingly obvious to me that the right teacher or method for some children may be wrong for others.[35] Even for the same children, different approaches may work at different times. Teachers should be more prepared to specify their styles to the situations at hand, and administrators should be more prepared to select teachers for the students they will have.[36] This means we should learn more about appropriate ways to deal with children starting from a knowledge and acceptance of their present receptivity.

On Statistical Inference

Perhaps more mileage has been implied from the crude statistical estimation than can legitimately be claimed. The F test for sameness of regression coefficients is sensitive to the range of the observations and the linearity assumption of the regression. I explicitly stated that I do not assume that linearity holds, though one could define an "average" effect which is the linear fit. By stratifying on social class variables, then including correlates of social class in the equation, the likelihood of the fit being subject to nonlinearities is particularly severe.[37] For example, picture a circle of radius 10, centered at (0,0) on conventional Cartesian coordinates. Consider the upper half of the circle as the shape of the relationship being investigated. Suppose the data for the entire sample runs from -10 to 4. Then we will find a positive slope coefficient for the range of the observations. Suppose we split the sample: from -10 to 0, and from 0 to 4. Then we will have a negative slope for the upper sample, a positive slope for the lower sample, and a positive (but lower) slope for the pooled sample. The test might say that these were samples from different populations. The truth is that the calculated average effect in the first place was a function of the range of observation (for the slope would have been 0 if -10 to +10 had been observed), that the population fitted the true relationship perfectly, but the F test says these are most likely two different populations being sampled.

This sounds harsh, but it is important to demystify the notion that involved statistical models can, of themselves, confirm or deny hypotheses. That whole procedure is involved with the nature of the data, the range of the observations, the amount of knowledge external to the data, the complexities of the relationships and the simplicity of the equations, etc. I will propose here how the tests conducted above might be amplified upon. I plan to investigate another city in the EEO data. I will code that city's data the same way, and test whether the middle class whites in that city and in Eastmet can be said to derive from the same population.[38] If the two white populations react the same way to school variables, but the black populations do not; if the middle classes do, but the lower and possibly upper classes do not; then the case will be quite a bit stronger. If all groups are unlike each other, then the test will say nothing, and one should feel dubious about the conclusions I am now drawing.

Teacher Specificity and Schools

Briefly, the argument of this paper has proceeded in this manner: Two methods of associating school resources with variation in cognitive outcomes (verbal, reading, and mathematics tests) were presented. Single linear regression estimates were derived for a single city, Eastmet, on observations of children who had not changed elementary schools, stratified by race. A three equation system with simultaneous estimation was also offered on this sample. The equations were compared between the races, and the associations between school variables and outcomes were found to be different. Some difference also was suggested between bottom quartile whites and the rest of the whites. An interpretation offered was that those school characteristics which affect whites, particularly middle class whites, are different from those characteristics which affect blacks and lower class whites. This was not the only possible interpretation, and indication was given of research in progress on this question.

Characteristics which are associated with outcome are called "resources," the amount of their "resourceness" to the different populations being indicated by the relative size of their coefficients. Teacher "specificity" then refers to the theory that certain characteristics have more resourceness for some children than for others. Since this concept is commonly accepted in the area of teaching exceptional children, an appendix reviews some of the special education literature (that dealing with integrating exceptional children into the normal classroom).

I argued that these concepts could be applied to situations in which not "normalcy," but simply differences among children in response to similar characteristics was the issue. Unfortunately, the literature on special education is not convincing about the nature of the characteristics of special teachers. "Empirical proof of the validity of special preparation does not exist. . . .Proof must be forthcoming that there is more special about special education than the children assigned to these classes." [(38), pp. 245, 246] Nor, in comparative studies, were the characteristics of either the teachers or the students in the special and the regular classes examined. Conflicting findings indicate to me that there might be some powerful variables at work which need to be investigated.[39]

One such type of variable might be a trainable teacher characteristic. If the evidence that there are teacher characteristics which affect output is considered weak, then the argument for specificity of this effect is equally weak, and the implication that such a characteristic is trainable is weaker still. Thus this paper is a tentative dip of the foot into the pond. The temperature feels right, but I would prefer to know about the temperament of the fish before actually advocating that we swim.

I am nonetheless willing to ask what swimming in this pond would be like, if the fish proved friendly. For that reason, I suggested that teacher specificity did not necessarily lead to segregated education, although most elementary education *is* segregated, and teacher training and hiring might therefore take note of those characteristics which are most useful for the particular children which the teacher will have. Teacher certification by *one* set of standards is perverse if teacher specificity has any validity at all. A highly verbal teacher might be such a resource that he might not need to fulfill other requirements, such as college graduation. Or perhaps some children need more attention paid to them than a single teacher can produce in a day: several part-time teachers might man one classroom. Perhaps some children learn best from "call and response" techniques, in which case a teacher with strong vocal chords and a room with soundproofing are resources.

These are just ideas. Some are being tested, others should be. Meanwhile, how ought schools to be structured? In the absence of answers, what do we do?

Inertia or Control?

The history of education, as any other public institution, is one of inertia. In the absence of information—though usually the impetus is a belief which may or may not hold true—a bureaucracy tends to make minimally disruptive decisions. And bureaucracy is the name of the education game. It takes an aroused public to stir the system, and the evidence presented here is not the kind to kindle the public spirit. I do not envision an enraged mob storming the educational portals, demanding "teacher specificity for all!"

Despairing of a revolution of the masses, I still plead for changes in the structure of decisionmaking (a revolution by another name).[40] Specifically, at first, for principal-power. I would like to see each principal given a *budget* from which he could purchase resources, instead of being sent inputs (which may not be resources) from the central board. For example, some schools ordinarily cannot get substitutes. Under the present structure, they do not get the salary of the substitute spent in their school unless it is spent on a teacher. The principal, in effect, has a coupon from the board of education which is redeemable only in teacher services. No teacher, no redemption. All I am advocating is that the nature of this coupon be expanded: it should be able to purchase *any* educational service. A television set, perhaps; but that is not very imaginative, and given the nature of most television programs, not very educational. Perhaps art materials with which the students could decorate the teacherless room.

I can lose the point by being too specific. The possibilities should not be limited to my imagination and inexperience. Nor should they be limited by our notion of principals as they are now. If most principals, unable to cope with such new responsibility, would make essentially the same decisions—hire the same teachers, purchase the same other inputs—as they do now, then what is lost? If *some* principals struck out into new forms of school organization, then what possible gains! Most importantly, the principal with the power to decide how his own school would operate would have to respond to the community, including the teachers. This has both the dangers of faddism and the possibilities of relevance about which we are all aware. At the moment, I am more impressed with the possibilities.[41]

Not just the ratio of teachers to other resources, but the type of teacher, should somehow be more a matter for local control, relating to the students. A principal might want to have one very expensive (but charismatic) teacher, and several community aides who are underpaid volunteers. Or he might want a teacher who is not acceptable to the school board, because that teacher has the specific talents needed in the school, but not the nominal qualifications. A principal might be restricted by his community from hiring unconventional teachers. But now he is restricted by his school board. And "unconventional" teachers is exactly what "teacher specificity" must mean. Eventually, if teachers appropriate to the situation are induced into schools, the conventions will change. Conventions are what schools of education transmit. So I contend that the place to start change is the public school, and the way to start is with principal control of his budget. Experimentation could take place within this context, and teacher specificity investigated. Then, with an idea of what kinds of things produce results for different kinds of children, teacher training can attempt to "produce" the kinds of teachers being called for.

Obviously such an idea as principal-power needs more exposition, more defense.[42] But so does the concept of teacher specificity. The two are somewhat tied together, though, in that the allocation decisions implied by teacher specificity seem too difficult for large central control. A central board might act as a referral agency, taking "want ads" from principals, and "personals"

from prospective teachers. But such decisionmaking as I envision, based on the school needs, must be local. The point of this ending, then, is merely to indicate some of the implications of such a seemingly technical idea as the association of teacher resourceness with children's characteristics. If that concept seems reasonable, then perhaps the places it leads will seem more reasonable now than they once did. That would be a happy outcome of a long article, one as difficult for me to write, I assure you, as it has been for you to read.

APPENDIX

The Exceptional Child Analogy

Given the concept of the "normal child," to whom public schools address their attention, there must be the "exceptional child" who falls outside the range of ability described by "normal." Mackie estimates that 10 percent of the school-age children are exceptional on the low end, and 2 percent on the high end. "A total of 35 percent of all exceptional children were enrolled in special education classes in 1966." [(29), p. 5.] But the distribution of aid to exceptional children is not uniform by type of exception. Thus 50 percent of the blind and deaf, 80 percent of the mentally retarded, but 12 percent of the emotionally disturbed and socially maladjusted are in special classes.

I cannot here go into detail about the problems of diagnosis of exception, or even the concept of "normal" itself—the dimensions of normality which may be missed by standard measures. In fact, the whole effort of this paper might be seen as directed against the concept of "normal" children. I will devote some space to outlining the literature about integrating exceptional children into normal classrooms. Teachers are trained in one of two ways: specialists who see only the exceptional child and his teacher, and ordinary teachers who accept exceptional children into their classrooms with some training on how to handle the situation.[43] The point of this appendix is to investigate the extent to which teacher specificity and integrated classrooms are in conflict. The analogy between the situation of the physically handicapped child and the variations which I find in the "normal" category is not exact, but may lead to some insight into the question.

Those resources which enable a blind or deaf child to be integrated into the classroom are presumably not directly applicable to the ordinary child. But the presence of the exceptional child may benefit the others, as well as himself.

> It has been found that the sighted children in the school not only gain some insight into the abilities of one blind person but that some less enthusiastic pupils are motivated to better achievement while learning with a blind companion. [(21), p. 133]

Though we might accept such a "finding" with skepticism, the process which could create it is obvious, and its verity is possible. Not the *presence* of exceptional children, but their *success and acceptance* by the teacher could produce such reactions.

> Because these [exceptional] children will eventually be required to achieve a satisfactory adjustment within a predominantly normal society, the experiences they have as children with this society are invaluable to them. Furthermore, normal children should be given an opportunity to understand, accept, and adjust to children with exceptionalities. [(17), p. 3]

A resource to the exceptional child could produce a resource to the other children in the same simultaneous sense that a resource to grade aspiration produces verbal or reading score, though it is not directly associated with verbal or reading score, in the system presented above. The possibility that teachers can be trained to handle the special problems of the poor and culturally deprived is taken as a premise for most of this discussion, though there is no direct evidence supporting it.

Academic Achievement

The research on the success of integration of handicapped children is inconsistent. One study reports success; another, failure. O'Connor and Connor [32] report that children in special classes for the very hard of hearing (losses above 60 db) performed better than those integrated into regular classes, even after special preparation. Jones [21] found that visually handicapped children could be integrated; Fouracre [11] has investigated ways in which regular teachers could be trained to help the visually handicapped; and Leshin [24] and Berry [3] have separately stressed that such training *must* be given, because there are not enough specialists available. Edgerton implies that efforts to integrate mentally retarded may be misplaced:

> What I am suggesting is this: there is unquestionably some intellectual minimum below which no one can fall and yet claim competent membership in any society. We would all agree, I think, that no one whose IQ is 20 or 30 or 40 could become fully competent in any society. I am suggesting that the threshold between incompetence and competence in any society is actually closer to 60 or 70. [(10), p. 86]

Johnson's position [19] is much the same. Sparks and Blackman, on the other hand, report for the *educable* mentally retarded (usually IQ 75-90), "children in regular classes almost invariably demonstrate academic achievement superior to that of special class children." [(38), p. 243.] However, they also report that most studies are characterized by a "lack of control of the teaching in the experimentation." [(38), p. 244] Vacc [43] reports achievement gains for emotionally disturbed children were greater from special classes than integrated classes.

The parallel between teaching these specialized cases and teaching the disadvantaged has been made before. Tannenbaum notes that it is "entirely appropriate to canvass specialists in special education for some points of relevance between their unique expertise and the needs of the socially disadvantaged." [(40), p. 2] Jordan, however, warns against such facile comparisons. He defines the concept "disadvantaged group," referring to "a particular, discernible physiological defect," [(22), p. 314] and offers several arguments why the problems of the disadvantaged group are different from those of the "disadvantaged."

Far be it from me to try to draw strong conclusions from such a literature. But whether in special classes or in ordinary classes, "Teachers of atypical children require special training above that required for normal children." [(36), p. 81] And if more children were seen as "atypical," then more special training would be necessary. Edmund W. Gordon [(13), p. 15] suggests that the failure of EEO to find association between teacher characteristics and student output might be due to the teachers' failure "to plan learning experiences that outweigh home influences." He suggests that one *could* train teachers toward that goal, but he offers no evidence that this is possible.

The EEO findings, of course, can be faulted on statistical grounds, but Gordon's point is still important.[44] He reviews the literature on differences between lower class and upper class children, concentrating on their motivation. He concludes that the *values* of the children are the same, but the feedback to middle and upper class children is more direct. They do not learn delayed gratification, in essence, but have immediate gratification. Perhaps teachers have to learn how to offer *important* rewards to lower class children, but do not have to do that for other children.[45] Whatever the answer, if little can be said about school organization from the literature on special education, at least this much seems true of teacher training: we do not know what differential skills are required to produce academic achievement in different types of children. And this ignorance must produce failure.

Social Outcomes

What can special education in integrated setting do for socialization? Thurstone's [42] 1959 study is most often cited as evidence that the educable mentally retarded tend to have more friends if they are in special classes than in integrated classes.

Sparks and Blackman, who reported achievement gains for these children from integrated classes, report social gains from the special classes. Carroll, however, claims the opposite. "The current investigation supported the hypothesis that EMR children in a segregated setting would show less improvement in self concept than would EMR children in a partially integrated setting over a period of one academic year." [(6), p. 97] Darrah reports that special classes for educable mentally retarded "do not produce more potentially constructive members of society." [(8), p. 523]

Johnson and Kirk, studying social segregation, found mentally deficient children rejected by their classmates, but not directly "because they did not learn as fast as other children, because they did not read, or because they could not achieve in the academic areas. They rejected the mentally handicapped child because of his behaviorisms," such as teasing, cheating in games, and physical aggression. "These . . . can be interpreted as compensations for frustrations resulting from failure in school situations in which they cannot compete." [(20), p. 87] Vacc found that emotionally disturbed children also tended to be rejected by their classmates, but he did not ask why [43]. He found that behavior gains (Behavior Rating Scale) were greater for emotionally disturbed children (in matched samples) who had spent a year in special classes than those who had been in integrated classes. But no mention was made of the amount (or lack) of teacher training in the integrated classrooms. That is, this finding is consistent with my position that there is a teacher characteristic which is more a resource for emotionally disturbed children than for normal children. Presumably the teacher of the special classes in the study reported by Vacc had more of this resource, whether it be an attitude or training or whatever. If it is training, then his achievement and behavior results need not hold in the situation where the integrated class teacher has special training.

Rucker, Howe, and Snider confirm that mentally retarded children are less acceptable socially to their classmates than normal children, this time in a junior high school sample. [35] They also test whether the social ratings of the retarded children would be higher in a nonacademic class than in an academic. The differences, stratified by sex, actually went the other way. However, again the question "why?" was not asked. Since the "nonacademic" class chosen for this test was physical education, the hypothesis of Johnson and Kirk that academic frustration leads the retarded child to physical aggression could easily explain the finding: where better than in physical education class can one be physically aggressive?

The Analogy Reconsidered

The literature on the retarded and disturbed child is even less clear about the benefits of integration than that on the blind or deaf child. But several things do seem important. First, there seems to be a teacher characteristic which is a resource to these children in producing both affective and cognitive outcomes. Second, it is conceivable that the failure of integration is due to the failure of the teacher of the integrated class to have this resource. If this is true, and if, as in the case of the physically disabled child, integration seemed preferable to separation (except for some special classes), then whatever of this analogy is acceptable points clearly to more evaluation of what characteristics of teachers are necessary to integrate various children into one class. On the other hand, the basis of the analogy is just that only in special education is differential teacher training by type of child recognized. It is not clear that anything more can be drawn from such an analogy to the problem of different backgrounds among "normal" students. But it is an area worth investigating.

Acknowledgments

My colleagues at CEPR, David K. Cohen, Herbert Gintis, Christopher S. Jencks, Martin Katzman, and Marshall S. Smith, have all contributed to the production of this paper. In addition, the influence of Gordon Gillies, Mildred Howe, and Carol Stewart should be noted. Much of this work should bear the joint authorship of Henry M. Levin, from whose initiative the study was undertaken, and in conjunction with whom it has continued. Extraordinary research assistance was provided by Polly Harold.

Footnotes

[1] I am not sure this possibility is actually as likely as, in warning against it, I must assume it is. If the social outcomes are disastrous, the test scores are likely to be poor also. In fact, to assume that students could be both extremely alienated and maximum performers goes too far. But since skills as measured by test scores and other social outcomes are not perfectly correlated, the warning is still in order. And the question of deciding on a method when it helps some people but not others, and yet must be imposed on none or all—which is the nature of tracking—points out the inadequacy of correlation as a substitute for value judgments.

[2] The data and models have been derived jointly by Levin and myself both concurrently at Stanford and Harvard, and in summer work together at Stanford. Randall D. Weiss has also contributed to the formulation and estimation of the simultaneous equations model. The first person singular is used in this paper to assign responsibility, not credit, to myself.

[3] (See [1], [7], [16], [18], [28], [37].

[4] Maximization of a complement of our output measure would suffice, if the complementarity were strictly linear.

[5] One might object that if schools tried to maximize different things, they would not do so with the same kinds of inputs, but would employ those best for the output. For example, trade schools do not hire verbally proficient, but manually proficient teachers. However, elementary schools are equipped by tradition more than by rational management, the maximization of various outputs taking place on location, not overtly on central direction.

[6] Variations in inputs do not correlate with either output when the other output is not accounted for.

[7] Katzman [23] shows, for example, that the outputs of different elementary schools in Boston are quite different. I infer from his findings that the aims of these schools differ, though Katzman does not agree that this inference should be drawn. Different goals of schools, and the different goals which the school has for different children, is a vital problem in this type of analysis.

[8] Strictly speaking, I need only have eliminated those who had not been there since the fourth grade, since I used only the later grade teachers. However, the questions in EOS did not allow this distinction.

[9] Those pupils who said they were black *and* something else (Puerto Rican or Mexican) were included due to a doding error.

[10] This description of the school is esentially adopted from Mackler [30].

[11] There may be great reason to believe this, and it may be true, but no direct *statistical inference* of this nature can be made.

[12] In practice, this distinction is of little importance. Sixth grade teachers are not different from fifth grade teachers. Teachers in the fourth grade who were not in that school when the children were in the fourth grade were not eliminated, implying the assumption that they replaced teachers like themselves. The extent and direction to which this is biased is unknown, though replacement of likes seems more probable in high turnover schools, less probable in low turnover schools, where the replacement may be considerably younger than the person replaced.

[13] From conversation, I understand that Christopher Jencks is experimenting with this weighting scheme.

[14] Preliminary investigation indicated little success with principals' personal variables anyway.

[15] Unfortunately, however, 80 percent of the total sample had six or more of the items, the median being between seven and eight. For the samples actually employed in the regression analysis, 85 percent of the whites (though only 36 percent of the blacks) had eight or nine items. Thus the index does not necessarily contain the precision implied by nine questions. If that item which the children with only eight do *not* have is the same item for most of these children, the index merely measures the presence or absence of that item.

[16] The SES index was created by weighting the listed father's occupation by the mean income for his occupation and presumed race (from the race of the child) from the 1960 Census of Population reports for the area of the sample.

[17] As this is just an anecdote, not much analysis is required. But I did ask if the class knew either the children or the teachers, i.e., *knew* the track of each class from some external information. I was assured this was not the case.

[18] Even sex: 1 percent of the pupils in the SMSA sample from which our data is drawn gave no sex. I am not sure that all children who did not know their sex—or, more likely, could not read the question—did not mark it. There might be another 1 percent who randomly marked, and therefore one-half percent who are incorrectly coded by six. This is not enough error, surely, to cause mistrust of that variable example of how even the simplest item contains some error.

[19] The median might have been lower under guessing, since the random selection distribution is skewed about the expected value. The median was in fact *higher* than the mean. The expected mean under guessing would be below 10 if some students did not finish the test.

[20] Such independent determinations would violate the very concepts of joint production which they are supposed to estimate. In determining average effects, the production of other outputs is not accounted for, as it would be in joint production estimation, nor is an index of the joint product assumed to be maximized.

[21] The black equations have similar standard errors to the white equations, but the black dependent variables have smaller variances. In terms of standard errors, then, the black equations are just as "good" as the white equations, and the difference in R^2 might be considered a difference in the *data*, not in the equations.

[22] I refer to the variables as "family size," though the question asked for number of people living in the same home, which may include non-family. Because of the lack of variations in the possessions index, as noted above, a great deal of social class variation is left to be accounted for by other variables.

[23] The correlation between teacher racial preference and discrepancy is -.60 in the white sample. For blacks, the correlation is only -.06. Teachers of whites, then apparently are more free to follow their preferences in regard to race of their students than teachers of blacks.

[24] Rosenthal and Jacobson [34], but see their critics, for example Thorndike [41].

[25] More detail about these indexes will appear in future publications.

[26] It should be pointed out that 33 of the 35 Eastmet city schools had both white and black pupils. The weighting of resources, but not access to some of the resources, varied by race.

[27] The reader is reminded to refer to Levin [27] for details on simultaneous equation systems.

[28] The process by which this works is not clear, especially if grades do not correlate well with test scores, which often seems the case. If I had data on grades, the information system could be specified and the model would be greatly improved.

[29] For this reason, T statistics are not given for the black coefficients.

[30] Resources which induce discipline might stifle curiosity or inventiveness, for example.

[31] In the ordinary single regressions, large coefficients in the meaning given in the text below were considered if the T values were 1 or greater, even though not significant by conventional standards.

[32] I am not concerned with observed variation in teacher test score, because the observed variation may not represent the potential variation. However, this exercise comes dangerously close to using the equation for purposes it cannot perform, estimation of *marginal* effect.

[33] I am using here means of the samples containing 1,599 blacks and 1,727 whites. This is a reduction from 4,505 students in Eastmet after elimination of those reporting no sex, those neither black nor white, and those with incomplete records (students but no teacher, for example). This sample includes the suburbs of Eastmet, which gives a broader range of scores than the city sample alone.

[34] We do know that some children are generally recognized as class leaders, but that "outgroups" sometimes have their own leaders. We do not know the extent to which this leadership influences outcomes of schooling.

[35] Levin [25] gives an example which makes this point so clearly that conventional standards and measures appear ridiculous: "If black schools and white schools have the same number of teachers with the same preparation and experience, the two sets of schools are considered to be equal according to conventional criteria. Now, what if all of the teachers have white racist views?" Such views might not hinder, say, mathematics teaching in the white schools: but they might make serious teaching in black schools impossible.

[36] In the current school organization one could say this is done already: the better teachers, who might be able to adapt to the poorer students, nonetheless get the better students. The plea that administrators optimally assign teachers is empty within the current incentive structure. Optimum for whom?

[37] Note that I did not stratify by variables explicitly entered into the equation. Social class was defined by father's occupation, which is not used as a variable. However, social class is so highly correlated with possessions, family size, and father's occupation, that following a technically correct procedure is no salvation.

[38] Actually, there are more powerful tests than the F test on squared residuals with which I intend to ask these questions.

[39] Yamamoto [44], for example, found creativty measures differing in teachers, and unrelated to their backgrounds. He could not associate this difference to output differences, but he noted that his output measures might not have been appropriate.

[40] I will assume a public school system basically of the type we now have. Such schemes as "voucher plans," in which people purchase education from private schools, require more consideration than I can given them here.

[41] For a lucid discussion of the extreme public, the extreme private, and intermediate forms of school organization, see Levin [26]. Levin urges that "The time is ripe to experiment with at least one of these plans. . ." (p. 37). But why just one? A really daring experimental approach would outline the major differences in plans, and experiment with several so their outcomes could be directly compared.

[42] This idea goes farther than that proposed by the First National City Bank, that "Title I and other monies could be put to good advantage by providing principals of disadvantaged schools with discretionary funds to use for whatever special purposes they themselves deem necessary to improve achievement levels in their schools" ([33], p. 33). My suggestion is in regard to the entire school budget (except the capital budget, though there, too, the principal and his community should have a great deal more authority).

[43] Stephens and Birch [39] outline three organization plans for dealing with special education of partially seeing students which probably applies to most exceptional children; full-time special class, resource teacher, itinerant teacher. In the latter two cases, however, the child is placed in a regular classroom for most of his instruction. See also Fouracre [11] for school organizations.

[44] See, for example, Bowles and Levin [4], or the work in the body of this paper.

[45] I am frightened, however, by the report that monetary incentives have been taken into the classroom. Hamblin *et al.* [15] report that a market is set up in class, tokens given for "good" behavior, which are redeemable for candy. What is frightening is the idea that markets are a *socially desirable* method of accomplishing production. Markets may be efficient in terms of physical resources, but there is no evidence that they have anything but deleterious effects on attitudes towards other people.

References

1. Armor, David J. "School and Family Effects on Black and White Achievement: A Re-Examination of the USOE Data," prepared for *On Equality of Educational Opportunity,* edited by Frederick Mosteller and Daniel P. Moynihan, Random House: 1970, (forthcoming).
2. Berg, Ivan. "Rich Man's Qualifications for Poor Man's Jobs," *Transaction,* Vol. 6, March 1969.
3. Berry, Charles Scott. "The Exceptional Child in Regular Classes," *Journal of Exceptional Children,* Vol. 3, 1936-37.
4. Bowles, Samuel S., and Levin, Henry M. "The Determinants of Scholastic Achievement: An Appraisal of Some Recent Findings," *Journal of Human Resources,* Vol. III, No. 1, Winter 1968.
5. Callahan, Raymond E. *Education and the Cult of Efficiency.* Chicago: University of Chicago Press, 1962.
6. Carroll, Anne Welch. "The Effects of Segregated and Partially Integrated School Programs on Self Concept and Academic Achievement of Educable Mental Retardates." *Journal for Exceptional Children,* Vol. 34, October 1967.
7. Coleman, James S., *et. al. Equality of Educational Opportunity.* Washington, D.C.: U. S. Government Printing Office, 1966.
8. Darrah, Joan. "Diagnostic Practices and Special Classes for the Educable Mentally Retarded—A Layman's Critical View," *Journal for Exceptional Children,* Vol. 33, April 1967.
9. Dreeben, Robert. *On What Is Learned in School.* Reading, Massachusetts: Addison—Wesley Publishing Company, 1968.
10. Edgarton, Robert B. "Anthropology and Mental Retardation: A Plea for the Comparative Study of Incompetence." *Behavioral Research in Mental Retardation.* Herbert J. Prehm, Leo A. Hamerlynck, James E. Crosson, Editors. Eugene, Oregon: School of Education, 1968.
11. Fouracre, Maurice H. *Helping the Visually Handicapped Child in a Regular Class,* Teachers College, Columbia University, 1957.
12. Gintis, Herbert. *Alienation and Power: Towards A Radical Welfare Economics.* Ph. D. Dissertation (Economics), Harvard University, 1969.
13. Gordon, Edmund W. "A View of the Target Population," in *Special Education and Programs for Disadvantaged Children and Youth,* Tannenbaum, Abraham (editor), Washington, D.C.: Council for Exceptional Children, NEA, 1968.

14. Greeley, Andrew M. and Peter H. Rossi. *The Education of Catholic Americans,* Chicago: Aldine, 1966.
15. Hamblin, Robert L., David Buckholdt, Donald Bushell, Desmond Ellis, Daniel Ferritor. "Changing the Game from 'Get the Teacher' to 'Learn' ", *Transaction,* January 1969.
16. Hanushek, Eric A. *The Education of Negroes and Whites,* unpublished doctoral dissertation, Department of Economics, M.I.T., 1968.
17. Haring, Norris G., George G. Stern, and William M. Cruickshank. *Attitudes of Educators toward Exceptional Children.* Syracuse: Syracuse University Press, 1958.
18. Jencks, Christopher. "The Coleman Report and the Conventional Wisdom", prepared for *On Equality of Educational Opportunity,* edited by Frederick Mosteller and Daniel P. Moynihan and published by Random House, 1970 (forthcoming).
19. Johnson, G. Orville. "The Mentally Handicapped," Appendix B to *Attitudes of Educators Toward Exceptional Children, by Haring, Stern, Cruickshank. Syracuse: Syracuse University Press, 1958.*
20. *, and Kirk, Samuel A. "Are Mentally-Handicapped Children Segregated in the Regular Grades?" Journal of Exceptional Children,* Vol. 16-17, June 1951.
21. Jones, John W. "Developments in Oregon's Program for Educating Blind Children," *Journal of Exceptional Children,* Vol. 19, 1952-53.
22. Jordan, Sidney. "The Disadvantaged Group: A Concept Applicable to the Handicapped." *The Journal of Psychology, Vol. 55, April, 1963.*
23. *Katzman, Martin T. "Distribution and Production in a Big City Elementary School System," Yale Economic Essays,* Vol. 8, No. 1, Spring 1968.
24. Leshin, George. *The Exceptional Child in the Regular Classroom,* University of Arizona, 1967.
25. Levin, Henry M. "The Case for Community Control of the Schools," 1969, paper delivered at University of California School of Education.
26. . "The Failure of Public Schools and the Free Market Remedy," *Urban Review,* Vol. 2, No. 7, June 1968.
27. . "A New Model of School Effectiveness." 1970, prepared for Office of Education, Bureau of Educational Personnel Development, Conference "How Do Teachers Make a Difference?"
28. . *Recruiting Teachers,* Charles E. Merrill (forthcoming, 1970).

29. Mackie, Remaine. *Special Education in the U.S.: Statistics 1948-1966* (New York: Teachers College Press, Columbia *University, 1969).*
30. Mackler, Bernard. "Grouping in the Ghetto," *Education and Urban Society,* Vol. 2, No. 1, November, 1969, pp. 80-96.
31. Michelson, Stephan. "Resource Allocation: Reflections On the Law and the Data," *Inequality in Education,* Cambridge, Mass.: Harvard Center for Law and Education, 1969.
32. O'Connor, Clarence D., and Connor, Leo E. "Deaf Children in Regular Classrooms," *Journal of Exceptional Children,* Vol. 27, May 1961.
33. *Public Education in New York City,* First National City Bank of New York, November, 1969.
34. Rosenthal, Robert, and Jacobson, Lenore. *Pygmalion in the Classroom,* New York: Hold, Rinehart and Winston, 1968.
35. Rucker, Chauncy N., Clifford E. Howe, Bill Snider. "The Participation of Retarded Children in Junior High Academic and Regular Classes," *Journal of Exceptional Children,* Vol. 35, No. 8, April 1969.
36. Sengstock, Wayne L. "Contributions of Programs for the Mentally Retarded," in *Special Education and Programs for Disadvantaged Children and Youth* (Abraham J. Tannenbaum, editor). Washington, D.C.: Council for Exceptional Children, NEA, 1968.
37. Smith, Marshall S. "The Coleman Report: The Basic Findings Reconsidered," January 1, 1970, prepared for *On Equality of Educational Opportunity,* edited by Frederick Mosteller and Daniel P. Moynihan and published by Random House. 1970 (forthcoming).
38. Sparks, Howard L. and Blackman, Leonard S. "What is Special About Special Education Revisited: the Mentally Retarded." *Journal for Exceptional Children,* January 1965.
39. Stephens, Thomas M. and Birch, Jack W. "Merits of Special Class, Resource, and Itinerant Plans for Teaching Partially Seeing Children." *Journal of Exceptional Children,* Vol. 35, February 1969.
40. Tannenbaum, Abraham J. (editor). *Special Education and Programs for Disadvantaged Children and Youth.* Washington, D.C.: Council for Exceptional Children, NEA, 1968.

Chapter 7

POLICY IMPLICATIONS AND FUTURE RESEARCH: A RESPONSE

Robert M. Gagne

In the preceding papers, various models of the system of education have been proposed in the attempt to illustrate and dramatize the variables involved in the process of analysis. In order to aid in the formulation of my comments, let me first present the kind of model I have in mind, which I believe is representative of those used by a number of the authors:

The Education Model

Input Variables

Fixed — genetic constitution

Proximal — opportunities for learning

Distal — home and community environment, school environment, teacher climate, instructional materials, library holdings, etc.

Process Variables

Proximal — those human actions which transform distal input variables into proximal inputs

Correlated — teacher characteristics, abilities, length of service, etc.

Output Variables

Proximal (or Criterion) –What are students able to do?

Correlated – standardized achievement tests and attitude measures

The major difficulty I have in interpreting the results of the reported studies in this publication is that they deal with distal or correlated measures, and fail to use proximal measures. This of course is not a criticism of the methods of analysis employed. I am also fully aware that investigators have made serious attempts to find and use the "best" measures available. Nevertheless, in the light of the model presented, these measures are not good enough. As a result, the studies often have the appearance of correlating one measure of "academic intelligence" with another.

Regarding input variables, we all agree that the "fixed" variable provided by genetic factors is difficult to measure, and must for the time being be taken into account in other ways. The variable of direct relevance to the problem is opportunities for learning, and one seldom encounters such a measure in studies of the sort which have been discussed. Instead, a variety of distal variables are employed, including such things as home and community environment, family economic status, type of school, and others.

It is generally recognized that some attention needs to be paid to process variables, those human actions which transform the raw materials of input into opportunities for learning. Educational researchers tend to be highly aware of the discrepancies which often occur between "instructional materials" and "what the teacher does with them." Seldom do we find, in such studies as these we are considering, measures of process which are direct, in the sense that they indicate the nature of teacher activities. Again in this area, there is frequent resort to correlated variables such as the amount of teacher education, length of service, kind of experience, or personal qualities.

Particular attention needs to be given to output variables, but very little has been said about them in these papers. Here again one must recognize that achievement measures as obtained from standardized tests (of "reading," numerical ability, or whatever) do not provide direct measures of what students are able to do. Instead, they are correlated measures, possessing many of the characteristics of intelligence tests.

Here is a quotation from an article by Husek (1969)[1], describing the accepted method of developing achievement tests:

"Let us examine a hypothetical, good social studies teacher. Our teacher has been taught to try to specify his teaching goals in terms of behavioral objectives, and he also agrees that his best hope of evaluating his students is in terms of objective tests. So, he constructs a test to give to his students and over a period of several years discards some items and rewrites others in line with the results of item analyses which he faithfully performs. It does not make too much difference what kind of item analysis he performs, but let us assume that he uses something which tells him how well his items discriminate between the high scorers and the low scorers on the total test. Let us also assume that our teacher is a good one and actually gets across much of what he hopes he is teaching.

With these assumptions, what kind of test is developed? The item analysis procedure, first of all, eliminates items that everyone completes correctly or which everyone misses. This will mean that in the long run, especially if the teacher is a good one, most items which are directly related to the teacher's objectives will be dropped from the test because they do not discriminate among the students. This should not be surprising, and it is certainly not new. Thirty years ago Lindquist was telling test constructors that the objectives of a course would not be good sources for discriminating items. The developing test will also tend to become more homogeneous: isolated items will tend to be dropped, and items picking up similar information will tend to be selected.

In fact, over a period of years, I think that our hypothetical social studies teacher is developing a good general mental abilities test with items focused on the social studies. This kind of test may not be the kind of test the teacher thinks he wants, but it is certainly the kind that will produce variability in the student test scores."

It should be noted that the procedure followed by the teacher, as described here, is basically the same as that used to develop standardized achievement tests. It is clear, therefore, that such tests do not provide a direct measure of output in terms of what students are able to do. They are correlated measures, which makes them forms of "intelligence tests." While the methods of development are somewhat different, it seems likely that many attitude scales possess essentially the same inherent defects as output measures.

On the whole, then, these studies tend to exhibit an unfortunate circularity, owing to the fact that they employ measures which are not valid as direct indicators of input, process, and output. Is it possible to design studies which break this vicious circle, and approach the problem more directly? I believe that this could be done. It is conceivable, but by no means easy.

The simplest and most straightforward study would be that between a direct proximal input measure, or measures, and a direct output measure. The direct input measure might be something like "amount of opportunity for learning in school," or alternatively, "time spent in active learning." A direct output measure would take the form of "time to achieve specified performance objectives," or alternatively, "breadth of knowledge of the subject-matters taught in school." I need not reemphasize that the latter measure would have to be designed so as to ignore such characteristics as difficulty of items, and otherwise would studiously avoid other kinds of distortion of measurement.

If one were able to carry out this kind of study, he should also be able to apportion variance among various "input" variables other than learning time itself, such as home environment, classroom climate, peer influences, and others. Further, it should then be possible to go on to study directly what I refer to as "process" variables–what actually does the teacher do which makes a difference, given that there is a difference found in the first place. Then, if one were interested in further followup, he could tackle the "correlated variables" such as teacher characteristics.

In summary, my own reactions to the correlational studies that are reported is that their credibility is very low. I draw almost no conclusions from them. If an administrator or policymaker asks the question, "What do teacher characteristics have to do with the outcomes of school learning," the answer should be–"We have no way of answering that question at present. First, we have no measures of learning outcome worthy of the name. Second, we have inadequate measures of input. And third, even if we had such measures the question about teacher characteristics should not be asked until we know better what processes the teacher is employing to insure learning."

Now, obviously, there are many problems to be solved if we are going to get the measures that we now lack. They will not be solved by increasing the number of schoolchildren in a sample, nor by increasing the complexity of our statistical analyses. They will be solved by tackling first problems first–by keeping in mind that what we want is an indication of the nature and quality of output; which means what students are able to do, and what kinds of choices of values they make.

Footnote

[1] **Husek, T. R. Different kinds of evaluation and their implications for test development. *Evaluation Comment,* 1969, *2* 8-10. (Center for the Study of Evaluation, University of California, Los Angeles.)**

Chapter 8

COMMENTS ON CONFERENCE

James S. Coleman

The papers presented in this conference gave, I believe, an excellent summary of the current work being carried out in between-school comparative analysis of student performance with cross sectional survey data. A number of conclusions can be drawn from the survey represented by the conference. First, even with the crude instruments of survey data, it is clear that variations in teachers' characteristics account for more variation in childrens' standardized performance in cognitive skills than do variations in any other characteristics of the school. It is evident also that one major aspect of variations in teacher effectiveness is variations in the teacher's verbal skills. This general determinant of teacher effectiveness is strong enough to be evident with even the crude methods of measurement used in these studies.

Second, it is clear that little useful information concerning the specific factors determining variations in teacher effectiveness will be obtained from the present data sources (e.g., Equality of Educational Opportunity Survey), or from similar sources. There are two directions that research must go if it is to be of serious benefit to policies concerning teachers. One is more direct observation of teacher classroom behavior, so that the input variables or stimulus variables (depending on whether one describes the system as an economist would or as a psychologist would) are more directly measured. This implies also associating in an analysis a student with his particular teacher, rather than (as in most of these studies) associating a student only with averages of teacher characteristics in his school. Another direction that such research should take is observation that measures student gain in performance, rather than level of performance. Most of the studies reported here measured only level of performance. This introduces enormous problems in separating effects due to student differences and effects due to teacher differences—problems that are greatly reduced by using longitudinal data on performance. Longitudinal studies in which the same students' gains in performance under two different teachers (over a span of 2 school years) would be especially valuable, because in this way, the student could serve as his own control in the study of teacher effects.

A third general conclusion I would draw from the conference is that research to be useful for policy related to teachers should be framed in the presence of those specific policy questions. Unless this is done, the research is likely to be irrelevant to policy. For example, some policy questions concern teacher selection, others concern teacher behavior in the classroom. It may be that the same research project cannot easily answer both kinds of questions. Furthermore, some characteristics of teachers are possibly important for learning may be inaccessible to usual modes of measurement. Unless the policy questions are known in design of the research, these characteristics may be neglected.

A fourth general point reinforced by the conference is that research results cannot substitute for policy, but can only be one of several inputs to policy. This tends to be obscured by the economist's formulation of research results in terms of cost effectiveness, or achievement output per dollar input. Such formulations are seductively appealing, but the fact remains that student achievement is only one of a number of considerations in teacher performance, and dollar cost is only one of the costs of a teacher to a school system.

Appendix A

CONTRIBUTORS

James S. Coleman – the principal author of *Equality of Educational Opportunity* (also known as the Coleman Report) is presently a professor at the Johns Hopkins University. He was formerly research associate with the Bureau of Applied Social Research, Columbia University, and then became the director of Simulmatics Corporation. A member of the American Sociology Association, his most recent books include *Adolescents and the Schools* and *Models of Change and Response Uncertainty.*

Robert M. Gagné – currently at the Department of Educational Research, School of Education, Florida State University, is also President-Elect of the American Educational Research Organization. Previously he was on the faculty at the University of California at Berkeley, and served on the Board of the Far West Regional Education Laboratory.

James W. Guthrie – is presently an Alfred North Whitehead Fellow at Harvard University, on leave from the University of California at Berkeley.

Eric Hanushek – is a captain in the U.S. Air Force, currently an assistant professor in economics at the U.S. Air Force Academy in Colorado,

Hanushek—Cont'd as well as a consultant for RAND Corporation. His Ph.D. from the Massachusetts Institute of Technology concerned "The Education of Negroes and Whites."

Henry Levin —as an associate professor of education and economics at Stanford University, has also been involved with ESEA title I Task Force and the study of decentralization of large city school systems in California. He was one of the major authors of *Schools and Inequality* conducted for the Urban Coalition and published numerous articles concerning the Equality of Educational Opportunity Survey. He is also author of *Recruiting Teachers For Large City School Systems,* an unpublished manuscript.

George Mayeske —is a social psychologist at the U.S. Office of Education. He completed graduate work in psychometrics and social psychology at the University of Illinois and has worked in the Federal Government for 5 years in personnel research. During the past 3 years he has been involved with operations research in the Office of Program Planning and Evaluation in the U.S. Office of Education. He has written a large number of monographs concerning the Equal Educational Opportunity Survey data—some of the results of which are reported in this publication.

Stephan Michelson —is a research associate at the Center for Educational Policy Research and lecturer at the Graduate School of Education, Harvard University. He recently published an article in the new publication of the Harvard Center for Law and Education, *Inequality in Education,* titled "Resource Allocation: Reflections On the Law and the Data."

Alexander M. Mood —recently served as Assistant Commissioner for the National Center of Educational

Mood—Cont'd	Statistics in the U.S. Office of Education. As Assistant Commissioner he carried out the Equality of Educational Opportunity Survey with the service of outside consultants and contractors. (This report is often referred to as the Coleman Report after James Coleman, the principal consultant). Dr. Mood is now the Director of the Public Research Organization, University of California at Irvine.

Appendix B

CONFERENCE PARTICIPANTS

"Do Teachers Make a Difference?" sponsored by Division of Assessment and Coordination, Bureau of Educational Personnel Development, U.S. Office of Education. February 4, 1970

Speakers

Dr. Alexander Mood
Director, Public Policy Research Organization
University of California at Irvine
Irvine, Calif.

Dr. James Guthrie
Alfred North Whitehead Fellow
Harvard University
Cambridge, Mass.

Dr. Henry Levin
Associate Professor
School of Education
Stanford University
Stanford, Calif.

Captain Eric Hanushek
Assistant Professor
DFE United States Air Force Academy
Colorado Springs, Colo.

Dr. George Mayeske
Research Psychologist
Office of Program Planning and Evaluation
U.S. Office of Education
Washington, D.C.

Dr. Stephan Michelson
Research Associate
Center for Educational Policy Research
Harvard University
Cambridge, Mass.

Dr. Robert M. Gagné
Department of Educational Research
School of Education
Florida State University

Discussants

Dr. Albert E. Beaton
Advisor, Statistics and Data Analysis
Educational Testing Service
Princeton, N.J. 08540

Dr. Paul Campbell
Director, Bureau of Educational Quality Assessment
Pennsylvania State Department of Public Instruction, Box 911
Harrisburg, Pa. 17126

Dr. James S. Coleman
Professor
Department of Social Relations
Johns Hopkins University
Baltimore, Md. 21218

Mrs. Margaret Labat
Principal
Garnet-Patterson Junior High School
Washington, D.C. 20000

Dr. Fred McDonald
Associate Dean
School of Education
New York University
New York, N.Y. 10003

Dr. Richard Snow
Assistant Professor of Education
School of Education
Stanford University
Stanford, Calif. 94305

Mr. Ronald Tyrrell
Patrick Henry Junior High School
11901 Durant Avenue
Cleveland, Ohio 44108

Dr. Finis Welch
National Bureau of Economic Research
261 Madison Avenue
New York, N.Y. 10016

Dr. Doxey A. Wilkerson
Chairman, Curriculum and Instruction
Ferkauf Graduate School
Yeshiva University
55 Fifth Avenue
New York, N.Y. 10003

Bureau of Educational Personnel Development–*Staff*

Dr. Don Davies
Associate Commissioner

Mr. Russell Wood
Deputy Associate Commissioner

Mrs. Iris Garfield
Director
Division of Assessment and Coordination

Mr. Peter Hartman
Program Specialist
Division of Assessment and Coordination

U. S. GOVERNMENT PRINTING OFFICE : 1970 O - 388-306